Imaging for stereoscopic displays

Christopher Jones

University College London

Submitted for the degree of Doctor of Philosophy of the University of London

Abstract

This thesis addresses the problem of calibrating a stereoscopic camera with a minimum of necessary post-processing. This is achieved through a two step procedure, the first step of which is a calibration of the sensors in rotation by means of laser diffraction, without attached lenses. The second step involves attaching the lenses and using a simplified conventional image-based calibration to determine the effects of motions of the optical centres due to lens focusing. Mounting considerations and long-term stability are also addressed.

This method enables the construction of a stereoscopic camera which requires no interpolative rectification, with the calibration maintaining accuracy over a range of focal distances. Such a camera is built and calibrated, and tested to demonstrate the validity of the predicted error estimates. This approach is shown to be effective in producing stereoscopic images for display which meet the requirements of the human visual system. A comparison of this approach with previously published methods is presented.

Some or all of the techniques described in this thesis may be incorporated into existing calibration schemes to improve the quality of the produced stereoscopic images. The improvements provided by a hardware calibration as described may be especially valuable in applications where maintaining full sensor resolution in the displayed image is desired.

Table of Contents

Tables

Figures

Acknowledgements

I would like to take this opportunity to acknowledge those people who have supported me over the duration of my PhD research.

I would like to thank Dr. Neil Stewart (now at Sentec), who supervised me on behalf of Sira for the first half of my research. His advice and suggestions in areas of planning and prioritising, imaging techniques and technical issues are greatly appreciated. My thanks also go to Dr. Andrew Crookell, who supervised me on behalf of Sira for the second half of my research. His assistance with elements of the lens mountings were very useful, and his comments on the style and content of the many drafts of this thesis have been valuable in producing the finished document.

I would like to thank Prof. Alan Smith, who supervised me on behalf of University College London throughout my research, for his varied suggestions and his focus on the strategic aspects of planning this project to completion. His comments on the thesis drafts are also appreciated.

Sharp Laboratories of Europe financially supported the PTP, and provided most of the equipment used in the construction of the finished camera system. Use of their laboratories in Oxford was also provided for a few weeks, and I would like to thank them for their support and direction. In particular my thanks go to Dr. David Montgomery, whose technical knowledge of stereoscopy and in-depth discussions on many related subjects have been invaluable in helping to refine many of the ideas and arguments presented in this thesis. His detailed comments on the drafts of this thesis are also appreciated.

Sira has provided office and laboratory space throughout my research, and has facilitated most aspects of the PTP. My thanks go to them for their support and flexibility, as well as for the frequent technical assistance from their staff and inclusion in the Intelligent Imaging Programme meetings and relevant Sira training courses. In addition to my Sira supervisors, I would like to thank Dr. John Gilby in particular for his guidance through many aspects of the PTP scheme, and for his technical assistance and advice in imaging and mechanical fields.

I would also like to thank University College London's Mullard Space Science Laboratory for their technical assistance in mounting techniques and their initial assessment of the potential for adapting consumer imaging hardware. In particular, I would like to thank Dr. Dave Walton for his technical advice, and his comments on the thesis direction during the MPhil transfer procedure.

My special thanks go to my wife, Jennifer Jones, for her constant support, understanding and good humour at home during my research and thesis writing, and also to my mother, Carole Jones, for her support and encouragement throughout.

This research was undertaken within the Postgraduate Training Partnership established between Sira Ltd and University College London. Postgraduate Training Partnerships are a joint initiative of the Department of Trade and Industry and the Engineering and Physical Sciences Research Council.

1 Introduction

This introductory chapter sets out the basis for the thesis as a whole. The chapter begins in 1.1 with a description of the thesis objectives, and summarises the hypothesis, aims and major achievements. 1.2 introduces the field of stereoscopy, while 1.3 shows some of the history of stereoscopic imaging. 1.4 describes the current state of stereoscopy, and 1.5 discusses some applications of current and future stereoscopic technology.

In 1.6, the properties of the human visual system in terms of stereoscopy are discussed, and their implications for building stereoscopic cameras lead to the formulation of criteria which can be used to evaluate such cameras.

Finally, section 1.7 describes the scope of the thesis, giving details of the core hardware and software to be used, while 1.8 provides a clear overview of the structure of both the work and this thesis document.

1.1 Thesis objectives

1.1.1 Hypothesis

It is possible to design and construct a stereoscopic camera system which produces images for display that satisfies the requirements of the human visual system for comfortable, fused stereoscopic viewing. This can be achieved without any knowledge of the particular images for display, and need not involve any interpolative post-processing of the individual colour images.

1.1.2 Aims

- To define the stereoscopic viewing requirements of the human visual system in terms of a digital image.

- To determine the relationship between image point movement and camera component movement.

- To devise a set of reliable camera calibration techniques which calibrate the camera system such that minimal correction of captured images is required, and for any correction to be calculated and non-interpolative.

- To design and construct a stereoscopic camera system within a mounting framework which maintains the camera calibration while allowing flexibility of use.

- To meet or exceed the standards of camera calibration available with existing interpolative methods.

1.1.3 Major achievements

- Interpretation of the requirements of the human visual system in terms of camera component movement and positioning accuracy

- Use of laser diffraction from a repetitively-structured digital image sensor to position the sensors such that their rotations can be set with respect to an external coordinate system, decoupling the effect of camera rotation and translation parameters

- Use of a simplified lens calibration scheme which is accurate and robust yet contains no coupled parameters

- Use of highly accurate and robust mounting techniques to set components in position and maintain camera calibration

- Calculation of a camera yaw offset technique to minimise the effects of radial lens distortion in a stereoscopic camera system

1.2 Stereoscopy

"Stereo" is derived from the Greek word stereos, meaning "solid"[1], and means "relating to space"[2]. Modern common usage of the word stereo is usually applied to sound, but it can also be used to describe images.

Our natural mode of vision is stereoscopic. People have two eyes, separated in adults by a horizontal distance of around 65mm[3], which see slightly different views of the world. The brain processes these differing images, and fuses as much of them as possible to present a

"cyclopean view"[4] whereby the two views are seen as one. The brain also uses the differences between the images formed by each eye to give a sense of depth to fused objects in the field of view. This process is called binocular stereopsis[5]. Stereopsis enables us to judge relative distances of objects, as well as to help distinguish objects camouflaged against a more distant background.

A normal photograph contains only one view of the world, and so does not convey stereoscopic depth. It can contain other depth cues, such as perspective, but viewing them does not "feel" three-dimensional. However, the effect of having two eyes can be recreated by taking two pictures, from cameras spaced similarly to our eyes. Viewing the appropriate image with each eye results in a natural sense of stereoscopic depth[6]. The combined left and right images constitute a stereoscopic image. Sometimes people talk about "3D pictures" instead, which is less correct but conveys a sense of the extra information included. Stereoscopic images are sometimes referred to as "2.5D"[7], because they have some three-dimensionality but not as much depth information as in a full 3D representation of a scene. For example, a stereoscopic representation of a scene contains information about objects' positions in all three dimensions, but only contains information about visible points and surfaces, unlike a full three-dimensional representation which contains information about all points and surfaces, not just those visible from a particular viewpoint.

Stereo photography is useful because it presents a more realistic and informative image of a scene to a viewer than a conventional monoscopic photograph would. Examples of specific applications are discussed later in this chapter.

restricted

access

Image removed due to third party copyright

Figure 2 - View-Master reels and viewer

This was the first home stereo system to use colour, and was widely used. Around the same time, the technology of polarisation filters was advancing, allowing stereo projectors using this technique to be constructed. Though stereo photography was essentially a novelty at this time, it was again growing in popularity. View-Master's position was further cemented when it acquired the right to use Disney characters in its slides, opening its use up as a children's toy, and Fisher-Price Inc. (owned by Mattel Inc.) still market it today.

1.3.2 The repopularisation of stereoscopy

In the 1950s, do-it-yourself stereo photography became popular. Until that time professionals produced most stereo photographs, which were reproduced and sold on to the general public. However, the development of Kodak's Kodachrome 35mm colour film enabled compact, user-friendly stereo cameras to be built. The Stereo Realist, shown in Figure 3, sold 130,000 stereo cameras in this period, and about that many again were made by other people.

Image removed due to third party copyright

Figure 3 - the Stereo Realist[8]

Unfortunately, this interest died out in the mid-1960s, and many of the stereo cameras in use today date back to those from the Realist era. However, two cameras may be joined together in order to take stereo images, or an ordinary camera may be modified by use of a beamsplitter to

produce stereo image pairs[9]. Alternatively, a single camera may take both images in the stereo pair, simply being translated between exposures, but this method only works for static scenes.

The first truly three-dimensional form of imaging was invented in 1947 by Dennis Gabor - the hologram. Holograms give a true 3D representation of an object, which allow the viewer to see the imaged object correctly from different angles[10]. In the early 1960s their manufacture was simplified by the invention of the laser, but holograms remain difficult to produce compared to stereo pictures. Another form of producing genuinely three-dimensional scenes is volumetric imaging[11], which involves the formation of an image in a volume rather than a plane. However, volumetric imaging systems can only produce transparent objects, and currently operate at resolutions too low for a photographic-quality display.

1.4 The current state of stereoscopy

1.4.1 Methods of viewing in stereo

Stereo images can be viewed in a number of ways. These vary in their geometrical arrangements, their comfort, and the scale on which the images can be presented.

The simplest form of stereo viewing requires no special equipment to be used, and is often called "freeviewing". Using this method, the images are simply displayed side-by-side. There are two ways to freeview; parallel (shown in Figure 4) and cross-eyed (shown in Figure 5). The side-by-side stereoscopic images in this thesis can be freeviewed, if displayed or printed so that the separation is less than the viewer's interocular distance.

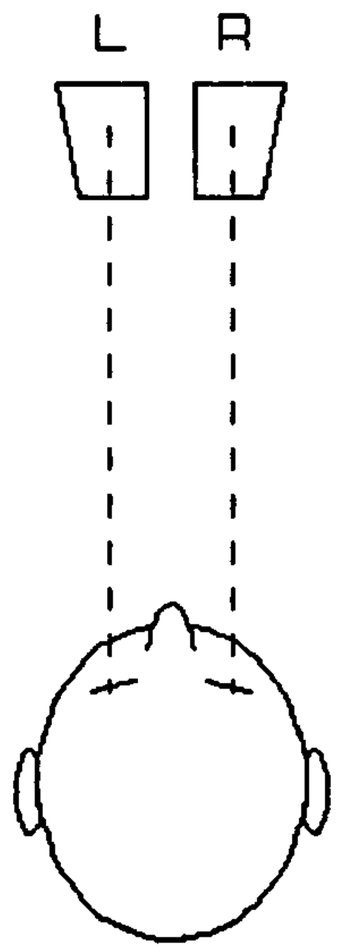

Figure 4 - parallel freeviewing

Parallel freeviewing is the most intuitive way to freeview. The left image is placed in front of the left eye, the right image in front of the right eye, and the eyes are allowed to relax until they converge at infinity. However, this does mean that images can't be any further apart than the interocular distance, since this would require the eyes to diverge in order for corresponding image points to fall on corresponding retinal locations. Though the technique is called parallel freeviewing, it is also possible to freeview with convergence at any point from the display plane to infinity, depending on the separation of the images. The essential component of "parallel" freeviewing is that the eyes converge behind the display plane.

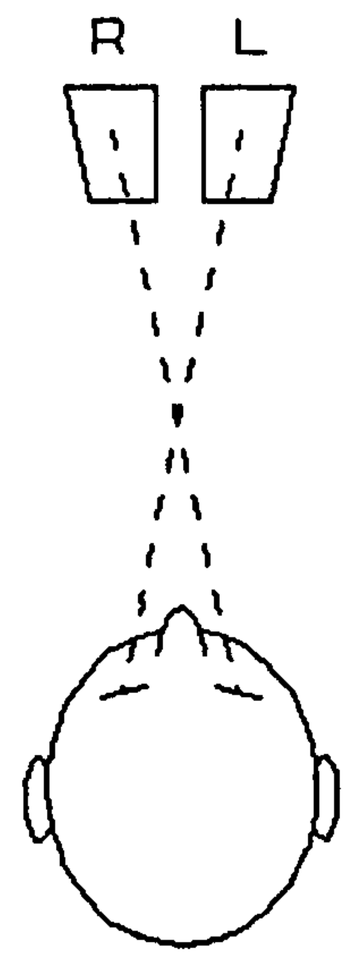

Figure 5 – cross-eyed freeviewing

Cross-eyed freeviewing is an alternative technique there the eyes converge on a point between the head and the images. In this instance, the images are mounted the opposite way round to

the eyes, so that the crossed technique will allow each eye to see the correct image. This technique is often found more difficult than parallel freeviewing, but allows larger images to be viewed. This is because with parallel viewing, the images can be no further apart than the observer's eyes, since the eyes do not easily diverge. In contrast with parallel freeviewing, images viewed using the cross-eyed technique can be much larger, since the separation is only limited by how much the observer's eyes can be crossed.

Freeviewing has the obvious advantage that it doesn't need a special viewer. On the other hand, the technique requires practice, and some people have a lot of difficulty doing it, while others cannot do it at all.

These difficulties might be expected to rule out freeviewing, yet it has already enjoyed huge popularity in certain applications, such as the "Magic Eye" random dot autostereograms[12]. These were made by taking a random dot image, and encoding depth information by horizontally shifting sets of dots by different amounts to create a second image. The two images can then be superimposed and freeviewed.

An alternative to freeviewing is the anaglyph[6], shown in Figure 6. This uses colour to communicate the stereo information, and relies on the viewer wearing coloured glasses. One channel is usually red, the other can be blue, green or cyan - there is no definitive standard.

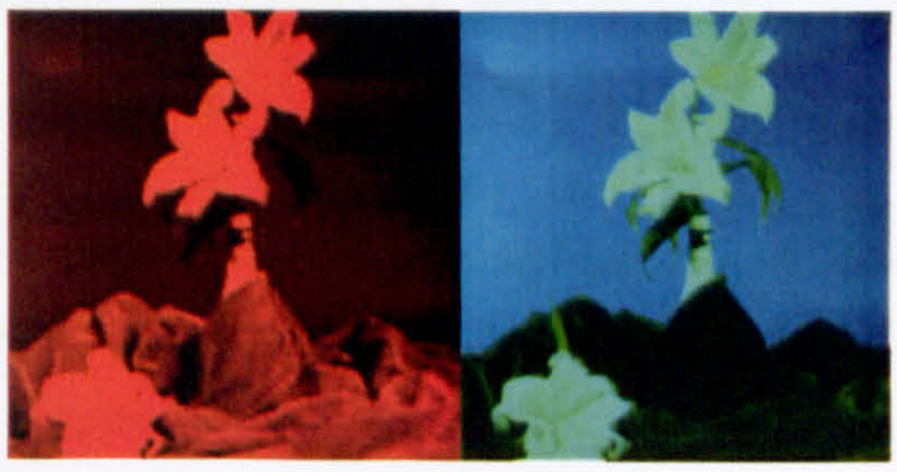

Figure 6 - anaglyph channels - red, and red removed

The two coloured images, each representing one stereo channel, are superimposed and viewed through colour filter glasses, so that each eye sees the appropriate image, as shown in Figure 7.

Figure 7 - combined anaglyph

Anaglyphs are comparatively easy to view, and cause relatively less eyestrain compared to freeviewing. However they do need a pair of glasses to produce the 3D effect, and there are associated problems. If an object in the original image is very red, for example, it will show up more strongly on the red channel than the other, and the stereo effect will suffer. If the filters used in the glasses don't completely remove the unwanted coloured image, then ghost images can be seen which cannot be stereoscopically fused, also known as crosstalk. People with some types of colour blindness, eye dominance and other visual problems can also experience problems with anaglyphs.

Despite their problems, anaglyphs are used extensively, particularly in magazines, because they can be viewed by anyone with the right glasses - which can be made of card and folded flat, and are inexpensive to produce.

Stereo can also be seen through slide viewing. This technique uses a pair of images with some kind of optical system that presents the correct image to each eye. These systems are very diverse, ranging from a simple pair of plastic lenses to a more complex piece of equipment such as a View-Master unit (as shown earlier in Figure 2).

This is a very good way for an individual to view good stereo images, from a quality and comfort point of view. Each eye sees the intended image with no eyestrain (at least none arising from the viewing process, the images themselves are another matter), and it remains popular among stereo enthusiasts.

Stereo images can also be viewed by field-sequential techniques, using shuttered glasses. In this method, alternate left and right channels are displayed at a fast frame rate, and the observer wears shuttering glasses that only allow the correct channel through at the correct time. The main drawback to this method is that the effective frame rate of the display is halved, which can be a problem with slow displays. Flickering can be a problem, as can image persistence effects in the display, which can cause crosstalk.

Another conventional method of viewing stereo is polarised projection. This is an excellent technique for presenting stereo to a large audience - the images are projected on to a silvered screen with differing polarisation, and the audience views them through appropriate polarisation filtering glasses. The screen must be of the correct type to maintain polarisation during projection - normal white screens are inadequate.

Though the polarisation method requires special projectors, screens and glasses, it remains unrivalled for presenting stereo images to more than one person at a time. There are fewer problems with ghosting than with anaglyphs, though the equipment is more costly.

In the last few years, development has proceeded on another type of stereoscopic display, which does not require the viewer to use any special eyewear. These are known as autostereoscopic displays[13], an example of which is shown in Figure 8, the structure of which is shown in Figure 9.

Image removed due to third party copyright

Figure 8 - an example autostereoscopic display[14]. The image on the screen is simulated.

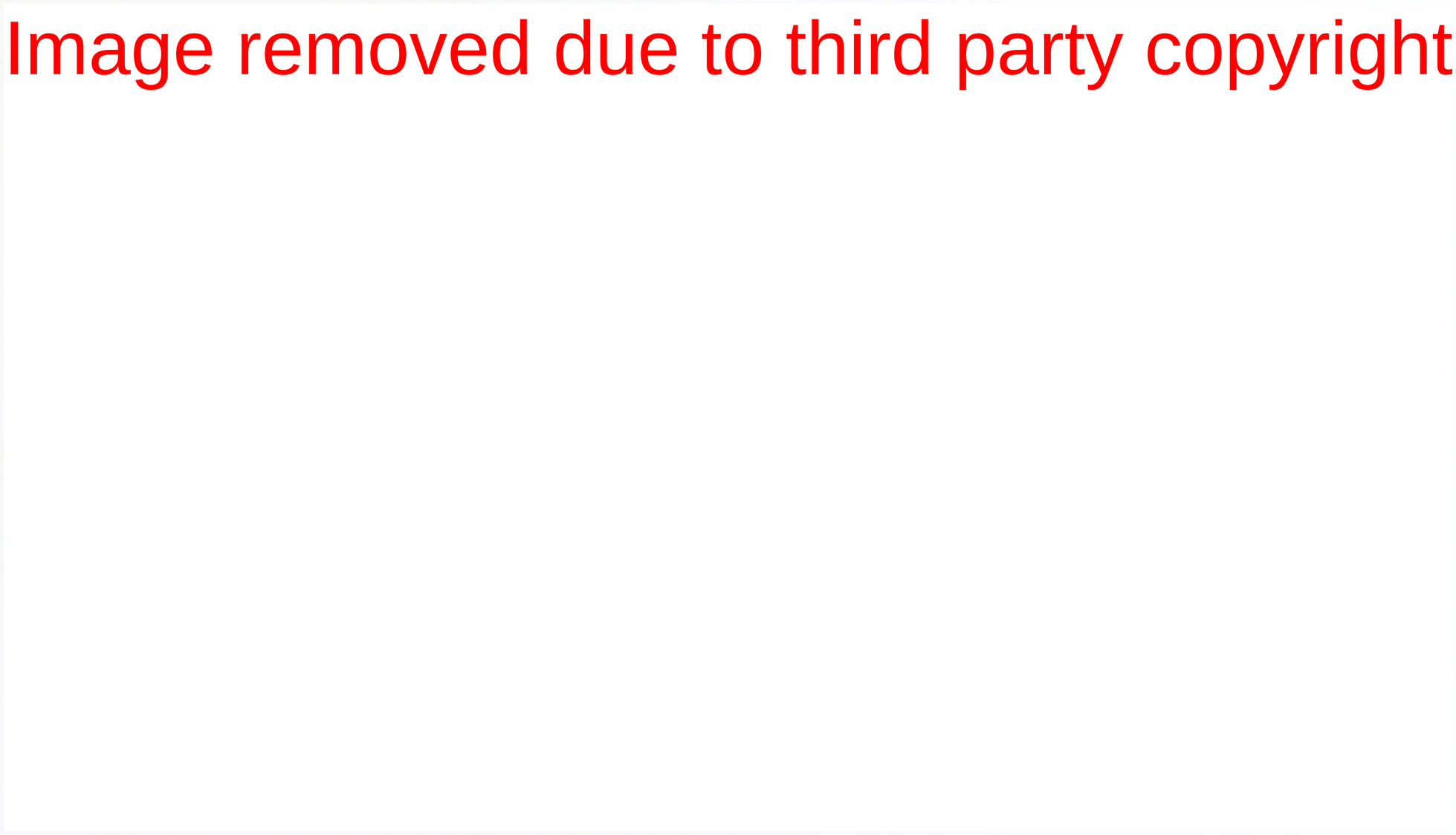

Figure 9 - autostereoscopic LCD[14]

The above figure shows how one kind of autostereoscopic display works. A normal LCD screen sits behind a parallax barrier element, which allows each eye to see only half the pixels on the LCD panel. For example, one eye sees the pixels marked as red in Figure 9, and the other eye sees the pixels marked as white. The barrier creates windows in the observer's space, which the observer must align with in order for each eye to see the correct set of pixels. The display may have a position indicator to aid this alignment. Head tracked displays are also under development, where the windows are moved by the display to follow the viewer's eyes.

A stereo effect can then be created by vertically interlacing the left and right channels on the LCD panel, so that the observer's left eye sees the left channel and the right eye sees the right channel. In some displays, the stereo effect can be switched off, restoring the full display resolution of the LCD[15].

These displays have the potential to become very popular, but there is a demand for a means of getting real world images for display. This need provides a large amount of the motivation for research into stereoscopic camera technology, and hence this project.

1.5 Applications of stereoscopic imaging

1.5.1 Examples

Stereo imaging has a variety of real-world applications. One example is photography, and many photographs can be enhanced by the use of stereo. A sample of such photography is shown in Figure 10.

Figure 10 – a stereo photograph, from the Ohio Stereo Photography Society

In addition, stereo imaging can be used for entertainment purposes, in videogames. Virtual Reality was a first step towards this, but today's sophisticated 3D games (an example of which is shown in Figure 11) should be readily adaptable to stereo display.

Figure 11 – Messiah, a 3D game[16]

Stereo can also be used in medical imaging, for example, to give doctors a clearer sense of the boundaries between parts of the body. It provides an alternative to false-colour enhancement, as the sensation of depth helps to distinguish objects from background. Stereo might be used in fields such as endoscopy, visualisation of scanning results and remote surgery via telepresence.

Architects and other CAD (Computer Aided Design) users can use stereoscopic systems to help create and present their designs, as even with computer generated images they generate a perception of solidity and reality. Stereoscopic displays can also be used to display a wide range of three-dimensional models, such as detailed molecular structures, which would appear cluttered or ambiguous in two dimensions.

Another potential application is in the remote operation of vehicles[17]. A sense of the three-dimensional structure of the field of view can be highly valuable, and a stereo imaging set-up would be especially useful to human operators in developing an intuitive feel for the vehicle's movement.

Other potential applications include visualisation of complex 3D data (such as air traffic control), remote sales and surveying.

1.6 Human stereo viewing capabilities

1.6.1 Overview

The human visual system has evolved mechanisms for processing images from two eyes, and presenting a single "cyclopean" image which is enhanced with, among other things, a sense of the relative positioning and distances of objects in a scene. These mechanisms are tuned to the images produced by the eyes[19], and to fool the brain into perceiving depth in artifical stereoscopic images, those images as presented by a display must be similar enough to ocular images to stimulate those mechanisms.

These requirements on stereoscopic images for display set requirements for images captured by stereoscopic cameras. The following sections describe the stereoscopic capabilities of the human visual system, and state the most critical image parameters for stereoscopic viewing. Later, in 4, these parameters are quantified in terms of the properties of a stereoscopic camera.

1.6.2 Physiology of the human visual system[19]

The structure of the human visual system is shown below in Figure 12:

Image removed due to third party copyright

Figure 12 - the human visual system[19]

The light-sensitive cells on each eyeball's retina feed information into the brain through the optic nerves. At the chiasma, the optic nerves meet and redistribute the nerves so that the nerve flow is routed according to the half of the visual field (left or right), rather than by eyeball. Each resulting optic tract contains information on only one half of the visual field, but contains fibres corresponding to both eyes. The optic tracts terminate in the lateral geniculate bodies, and signals from the eyes are loosely mapped, by the spatial direction they correspond to in the visual field, into the optic radiations[19].

From there, the signals proceed to the visual cortex, a single structure which receives input and feeds back into both hemispheres of the brain, and it is in this area in which various tuned substructures process, among other things, stereoscopic information[19].

This means that it is the properties of the visual cortex which define the stereoscopic capabilities of human vision.

1.6.3 Stereopsis and depth perception

It is important to distinguish between stereopsis and depth perception. When we look at the natural world, we have a sense of how far away things are - this is depth perception. This process operates via both monocular and binocular cues, while stereopsis operates as a purely binocular process[18].

In an ordinary 2D picture or video, we can have a sense of depth in the image. Cues for this perception include relative sizes of objects, interposition and occlusion, transparency, atmospheric perspective (illustrated in Figure 13), texture gradients, geometrical perspective, motion parallax, relative velocities and motion blur (the final 3 apply to video only). However, the perceived image retains a flat quality.

Figure 13 - atmospheric perspective - the mist blurs distant objects more than close ones

Stereopsis is the interpretation of depth through retinal disparities arising from the separation of our eyes. When we look at a three-dimensional object, the image of that object falls in a different relative location on each retina - a retinal disparity. When the eyes converge at a point in space, there is a locus of points that produce corresponding images when viewed, called the horopter[19,24]. Points falling in front of or behind this horopter show what are known as either crossed or uncrossed disparities, as illustrated in Figure 14.

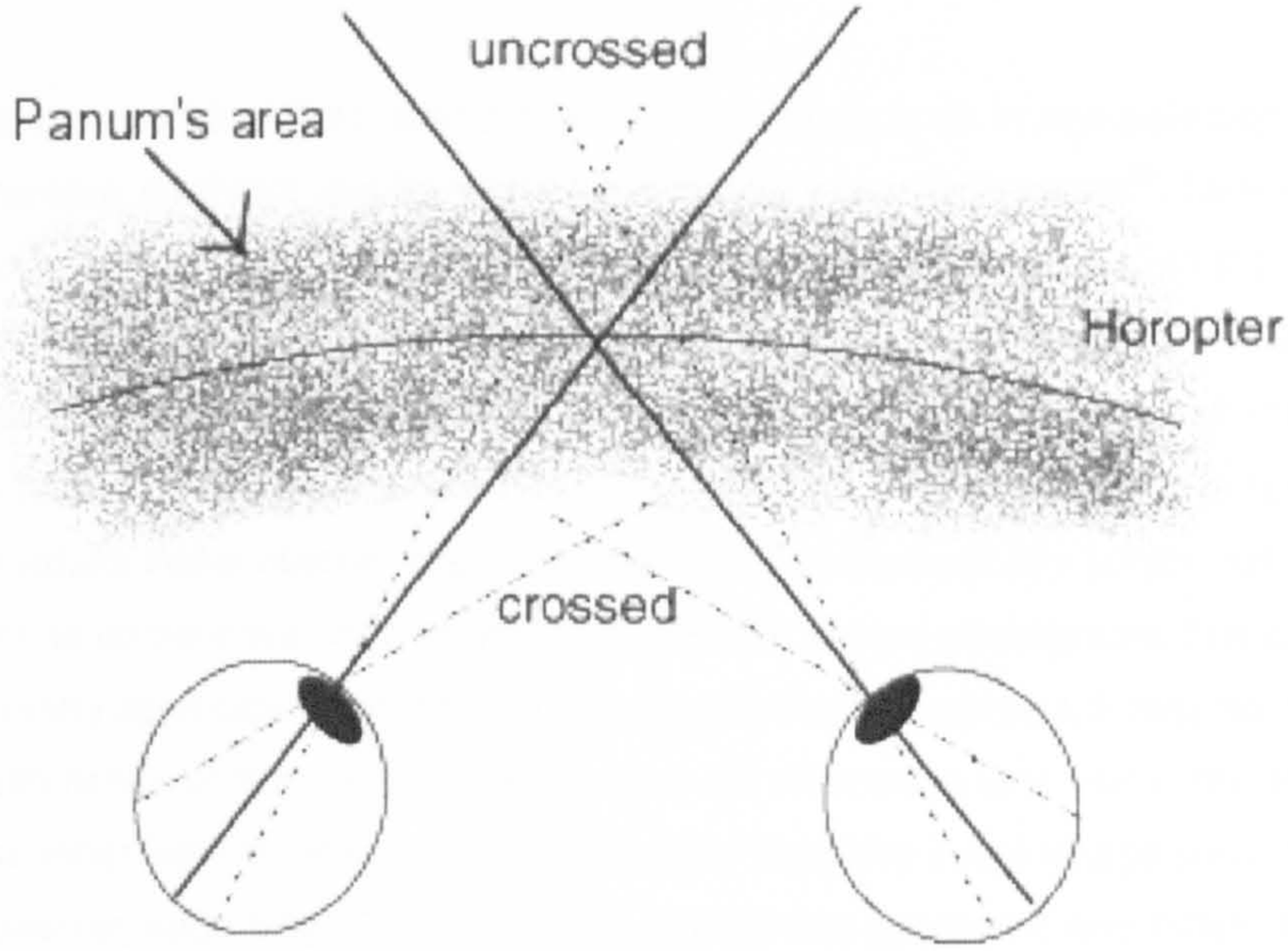

Figure 14 - the horopter and spatial Panum's area

For a given corresponding image point on the retina, an area around it will trigger stereo fusion although the images are not exactly corresponding. This is known as Panum's area[20,24], and extends approximately equally in both the horizontal and vertical directions[19]. This has an analogous concept in object space, and means that points within a certain distance of the horopter can be fused.

In normal, real vision, the restriction of stereopsis to Panum's area isn't a problem. When we look at something, our eyes focus and converge on it[21,24], and move the horopter so that the object falls within Panum's area and is in sharp focus. But in a stereo photograph, the camera has fixed the convergence and focus for the image, so the eyes can no longer move the horopter too far away from the display plane without losing stereo fusion.

1.6.4 Disparity

In a stereoscopic image, differences in the locations of a point's image in the individual left and right images, as seen by the eye, are called disparities. A disparity indicates to the brain a point's distance from the horopter, and generally smaller fusable disparities are more comfortable to view.

It is clearly desirable to be able to restrict the disparity range in an image pair captured by a stereoscopic camera so that it can be viewed comfortably by most people[22]. One way to achieve this is to narrow the separation between the two images captured, so that disparities are reduced correspondingly. Unfortunately, changing the stereo base can alter the perceived image, because the brain interprets disparities with respect to the separation of the eyes. A smaller stereo base (termed hypostereo[23] – a larger than normal stereo base is termed hyperstereo[23]) would make objects appear to the brain to be physically larger than normal, a technique which is sometimes used to dramatic effect in stereo photography but which is unsuitable for many applications. If the stereo base cannot be reduced, it may be necessary to restrict the depth range of the scene to be imaged. An alternative is to use a shorter focal length, giving a wider field of view, which reduces the disparity in the image plane for a given separation. However, wide angle lenses tend to suffer from geometric and other distortions more than narrower field lenses. These three methods for altering disparity can be balanced to give the best result for a particular scene.

Stereo viewing can be uncomfortable for people who are unused to viewing stereo images, because it may require the visual system to do something it is not used to doing. In everyday life, our eyes converge and accommodate (focus) at the same distance[24], so that whatever we're looking at appears sharp. In stereo viewing, the optimum focal distance is not always the same as the convergence distance, so the viewer may be required to decouple these two mechanisms (the alternative being to fuse objects not at optimal focus), and doing so usually becomes easier with practice[23].

1.6.5 Horizontal and vertical disparity

A point in the real world carries 3D information by virtue of the different image location it assumes when imaged from different points. An object will appear in different places in a pair of stereo images. This displacement can be resolved into horizontal and vertical disparities.

Horizontal disparities are often desirable in stereo images, since they are the source of depth for the eyes to interpret. Our eyes are horizontally separated, so we expect an image shift in the

horizontal direction between them, and our eyes are designed to converge in a horizontal direction. The horizontal disparity should not be too large (greater than the eye separation on the display, because most people's eyes do not readily diverge[23]) or stereopsis will fail, as image points lose correspondence on the retinas. Errors in horizontal disparity produce errors in stereoscopic depth but do not inherently cause problems fusing the images, provided the disparities remain within the limits described above.

Vertical disparities in stereo images are generally undesirable. Our eyes do not have a vertical separation (in our head frame - we can tilt our heads as we like), and so we find little vertical disparity while viewing the natural world. Unnatural vertical disparity in images is not well tolerated by the visual system, since it corresponds to a parallax in a direction that our eyes are not separated in, and cannot ordinarily converge in. Though stereopsis can be preserved for small amounts of vertical disparity, it is increasingly uncomfortable to view images with increasing amounts of vertical disparity[69]. For this reason, the minimisation of vertical disparity is a very important step in the creation of comfortably viewable stereo images.

1.6.6 Human stereoscopic vision studies

From studies of human stereoscopic vision, the average radius of Panum's fusional area has been determined to be approximately 6 arcminutes[20,24]. The way in which this translates to a size in an image depends on the conditions under which the image is viewed.

For an observer at a distance of 50cm from a digital screen, for example a computer monitor, with a pixel pitch of 0.26mm, 1 arcminute is equal to 0.56 pixels on the screen. So, Panum's fusional area has a radius of 3.4 pixels on screen, under those conditions.

Since binocular stereopsis requires that points to be fused fall within Panum's fusional area, and the eyes cannot converge in a vertical direction, stereopsis requires corresponding points to have vertical disparities within ±3.4 pixels across a stereoscopic image pair, for the viewing conditions described above. Viewing at a closer distance to the screen or using a screen with a higher pixel pitch will decrease the allowable pixel disparity, and vice versa. Horizontal disparity is not bound to be within the ±3.4 pixel limit, since the eyes can converge in the horizontal direction to allow corresponding points with larger disparities to fall within Panum's area.

1.7 Structure of the project

1.7.1 Objectives

This thesis is primarily concerned with the hardware aspects of building a digital stereoscopic still camera. Some hardware problems may be corrected for or alleviated by post-processing, but such processing frequently degrades the images concerned. On the simplest level, a sub-pixel shift or image scaling (with the exception of stretching by an integer factor) reduces the amount of original image information through necessary interpolations, seen as a slight blurring. In a more complex form of stereoscopic processing, disparity reduction algorithms may leave blank spaces in an image where occlusions have obscured the parts of the scene necessary to reconstruct that area. Simple croppings leave the bulk of the pixel data intact, but lose those parts that are cropped out. Not only does post-processing affect the quality of the images, it also requires processing power, memory, and time to achieve. It is clearly desirable to reduce the amount of processing to the absolute minimum necessary to produce correct and comfortably-viewable images.

The objectives of this thesis can be summarised as aiming to understand what is required to manufacture a high quality stereoscopic camera for producing stereoscopic images for human viewing, with minimal computer post-processing, and to produce such a camera.

1.7.2 Equipment considerations

The camera should have two heads, each containing its own optical system and sensor. There are many possible configurations, using either single or dual lenses and single or dual sensors, but the two-head system offers the most potential in that its drawbacks are more easily solved[25]. Two-head cameras are the most adaptable and versatile solution, are easily synchronised and can easily operate in a monoscopic mode.

The camera should be digital, for a number of reasons. Using digital detectors rather than photographic film provides other advantages such as a consistent detector and cell size and a higher degree of flatness[26,27]. Flatness is particularly important since out-of-plane distortion of photographic film frequently produces significant distortions in captured images. Consistency is also desirable because calibrations will be able to apply to a specific physical configuration, rather than a sequence of configurations which change every time the film moves.

The camera's axes should be parallel rather than converged, to eliminate keystone distortion and depth plane curvature[25,69] due to convergence. The only conventional advantage of

converging axes is that such a configuration is useful for imaging close objects when the camera separation cannot be reduced enough to keep the horizontal disparities within acceptable limits without excessive cropping. With a parallel axis configuration, other solutions can be found, such as restricting the depth range of the scene or using wider angle lenses, which do not produce such adverse image distortions. Such a configuration can still produce a 90% image overlap, using an 8mm lens, 2/3" format CCD, a camera separation of 10cm and an object distance of 1m. A parallel configuration also simplifies the design of the camera's mounting, since it would not be required to accommodate varying convergences.

The camera should be able to be adjusted in its axis separation. This allows for maximum flexibility in capturing images, since the range of horizontal disparities can be controlled.

The camera will not make use of zoom lenses. This is partly because zoom lenses are generally not controllable or reliable to the same degree that most non-zoom lenses are, because of the requirement that they are able to adjust their focal length. While these problems may be soluble, it is more important to solve the other problems first. Since zoom lenses are an optional part of creating a stereo camera, zoom lenses are not be examined as part of this thesis. The lenses used should be broadly similar to those used in general consumer 35mm photography in terms of field of view, with as little distortion as possible. Low distortion is desirable because lens distortion correction relies on interpolating software, which, as it has already been explained, is undesirable.

The size and appearance of the camera system are unimportant at this stage. While a compact and attractive camera is undoubtedly desirable for many applications, in this early design it is considered more important to understand how to produce as functional and accurate a camera as possible.

1.7.3 Camera hardware

The cameras used are Basler A101CPs[28] (formerly A113CP). These are 2/3" format CCD cameras, with a 1300x1030 pixel detector chip having 6.7μm square pixels, and a Bayer RGB colour filter array[29]. These cameras are selected for their small size (in order to achieve small separations in a parallel configuration), high resolution (in order to capture the most information possible within a scene), accurate exposure control (in order to take both images at precisely the same time), and progressive scan output (to give a high-quality image free from interlaced effects). They connect to the host computer through Imaging Technology PC-DIG framegrabbers, which are chosen because of their compatibility with the Common Vision Blox control tools, for ease of programming, and their ability to accept an external trigger, for purposes of synchronisation. The cameras are triggered using an Imasys Imatrig trigger unit, which allows the synchronisation of the two cameras. The main lenses are Schneider Cinegon

8mm lenses[30], which are optimised to have low radial distortion (3% barrel distortion). Low distortion lenses produce lower stereoscopic errors due to their distortion, and are the appropriate choice for a fully hardware solution. The lenses attach to the camera using a C-mount.

1.7.4 Software

Basler provide their own camera control software in the from of a simple GUI. In addition, the cameras may be controlled using basic serial port communication through an application such as Hyperterminal[31]. Various applications are used to interpret and display the data from the framegrabbers, mostly written using the Common Vision Blox[32] tools. The primary stereo camera GUI is self-written in Microsoft Visual C++ and is used to display and capture simultaneous images from the camera pair.

1.8 Overview

The work in this thesis, and its organisation, can be summarised in Figure 15 below:

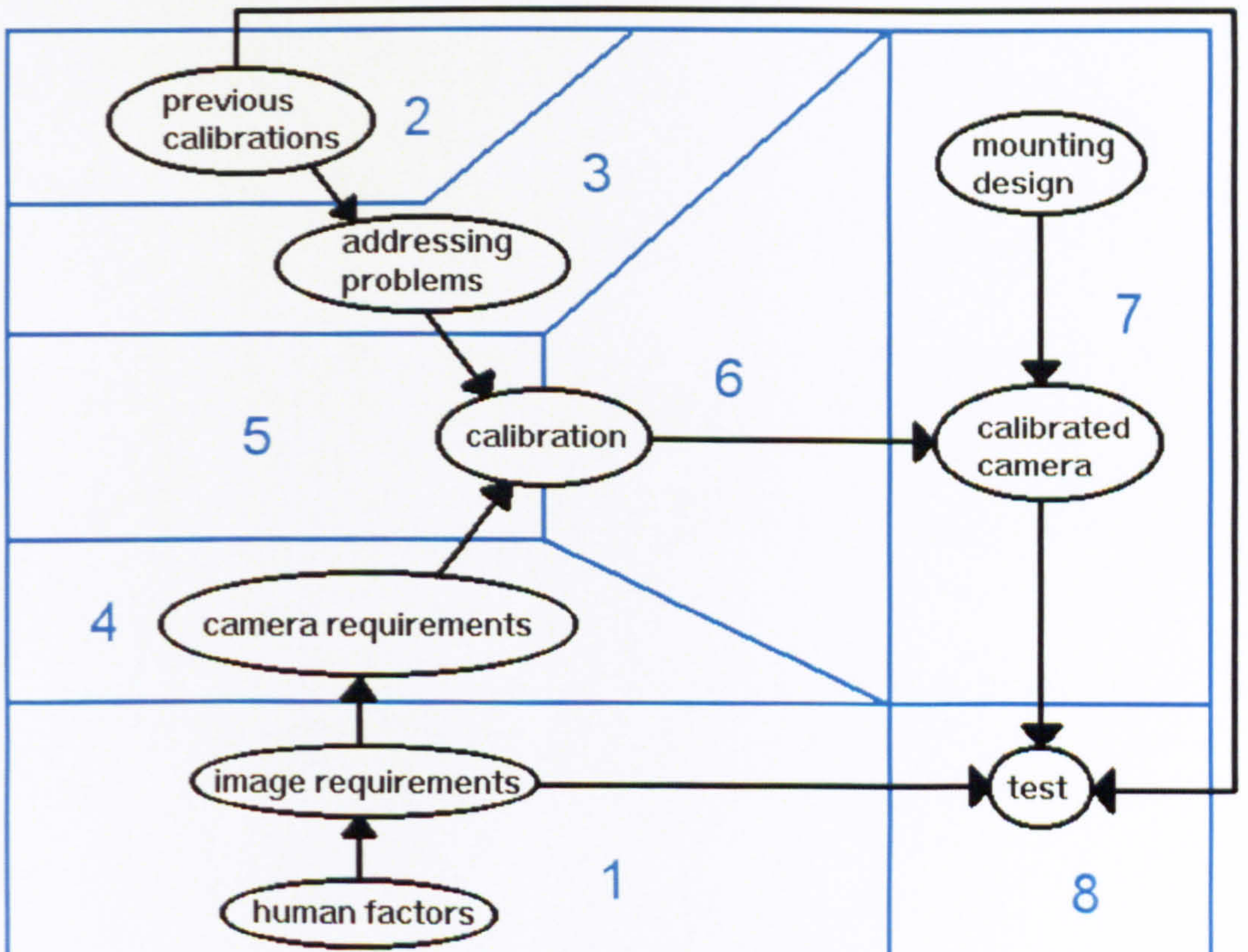

Figure 15 - thesis overview. Chapters are designated by the blue divisions and numbering

The properties of the human visual system set the requirements for stereoscopic images, as described in 1. These can be related to requirements for stereoscopic cameras, this relationship is calculated in 4 and is used to define the accuracy of the proposed camera calibration technique, described in 5 and 6.

The past work in the area of camera calibration, summarised in 2, is built upon in devising an improved calibration technique, and the basis for improving past techniques is set out in 3. These improvements lead to the description of the new calibration technique, described in 5 and 6.

The new calibration technique described in 5 and 6 is used, along with mounting techniques described in 7, to build a calibrated stereoscopic camera, also described in 7. Images taken with this camera are evaluated in the context of both past calibration techniques and the requirements set on the images, in chapter 8.

At the end of the thesis, 9 models the effects of lens distortion, and describes a new method for minimising those effects through stereoscopic camera calibration. 10 draws conclusions about the work contained in this thesis, and describes suggestions for further work and possible applications. Finally, 11 provides concise step-by-step descriptions of the calibration process, and 12 contains the list of references cited in this thesis.

2 Background

This chapter summarises previous relevant work in the area of stereoscopic camera design and calibration. 2.1 discusses digital cameras, and the stereoscopic implications of using charge coupled devices for image capture. Stereoscopic camera calibration in perhaps its conceptually simplest form consists of calibrating two single cameras, and 2.2 describes the available schemes for this. 2.3 describes the four basic stereoscopic camera design types, and their relative merits, while 2.4 discusses issues in and available schemes for stereoscopic camera calibration. Finally, 2.5 evaluates the existing material and identifies the deficiencies to be addressed in this thesis.

2.1 Monoscopic camera technology

2.1.1 Digital cameras

Traditionally, photography has used film-based cameras. These use a light-sensitive film, which is exposed to the scene for a short time, and later developed and processed to produce a printed picture. This method has the advantage of low cost but is restricted by the need to keep replacing the film, and the time-consuming developing process. Some cameras such as the Polaroids can produce nearly instant pictures, but the quality is generally poorer.

Digital cameras need no film replacement or developing time, instead capturing the images on a light-sensitive electronic sensor. The pictures are ready for transfer and display as soon as they are taken, and are also ideally suited to the growing use of home editing, since they can be manipulated on a home computer. Images can also be inspected and selected for retention or erasure immediately after exposure.

Digital cameras are a recent entry into the mass market, the first true example arriving in late 1992. The Logitech Fotoman[33], pictured in Figure 16, was a simple black and white camera, with 376x284 pixels, but was followed by a rapid flow of more and more sophisticated cameras.

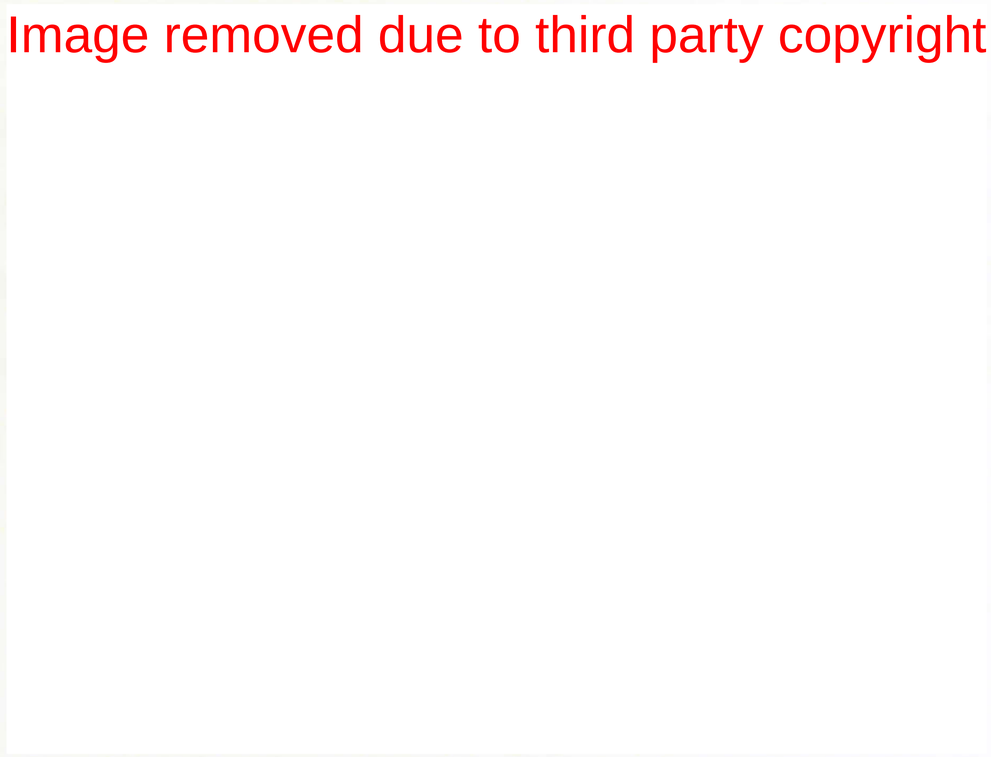

Figure 16 - the Logitech Fotoman

Today, digital cameras are of a quality that compares well to film. At the time of writing, one of the most advanced is the Nikon D1X, shown in Figure 17, which has a 4024x1324 pixel sensor, outputting 12 bits per channel RGB colour.

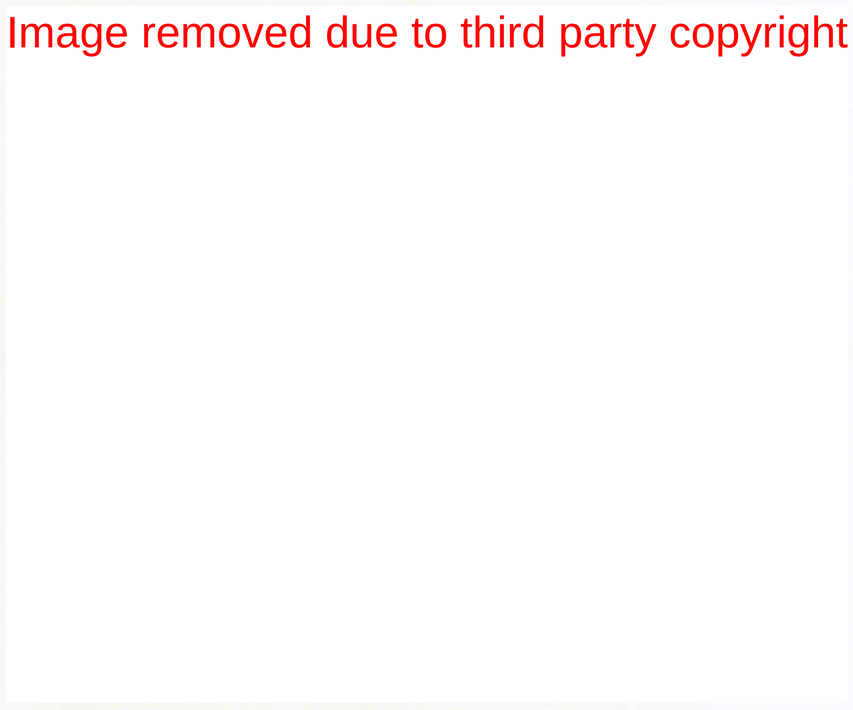

Figure 17 - the Nikon D1X

Digital cameras also have the ability to carry out image processing within the camera itself, which is important for many applications. Instead of applying special colour filters to the camera, the colour balance can be changed internally, for example. Overall, there is a high degree of flexibility in the use of digital cameras, which makes them ideal for use in many demanding applications. They are also well suited to being used to capture stereo images for display on a digital screen.

2.1.2 CCDs

Digital cameras use photosensitive silicon sensors. Currently the highest quality sensors are Charge Coupled Devices[34,35] (CCDs), though in the future these may be replaced by CMOS[36] sensors capable of on-chip image processing. CMOS sensors are cheaper and easy to produce, but CCDs may have the advantage of not requiring long exposure times (which is

important for imaging where the camera and objects in the scene are moving relative to each other), partly since the bulk of the chip can be light sensitive. This sensitivity helps to give higher signal to noise ratios, and for this reason they are most suited to high-quality imaging. However, they have their limitations, the most relevant of which to stereoscopic imaging are discussed in sections 2.1.4 to 2.1.11.

2.1.3 CCD operation

When a CCD is exposed, incident photons cause electrons to accumulate in the pixels. These charge packets are passed across the sensor and transferred in series to electronics which measure the magnitude of the charge. The transfer or charge from one pixel to the next is achieved by means of charge coupling[37], illustrated in Figure 18 and Figure 19.

Figure 18 - how charge is moved from pixel to pixel by charge coupling[38], the red lines indicate the voltage applied to the pixels and the blue blobs indicate the position of collected charge

The charges accumulate in packets in the image array, held in place by the applied voltage. During charge coupling, the applied voltage is shifted across the array, and the charge packets are drawn across with it.

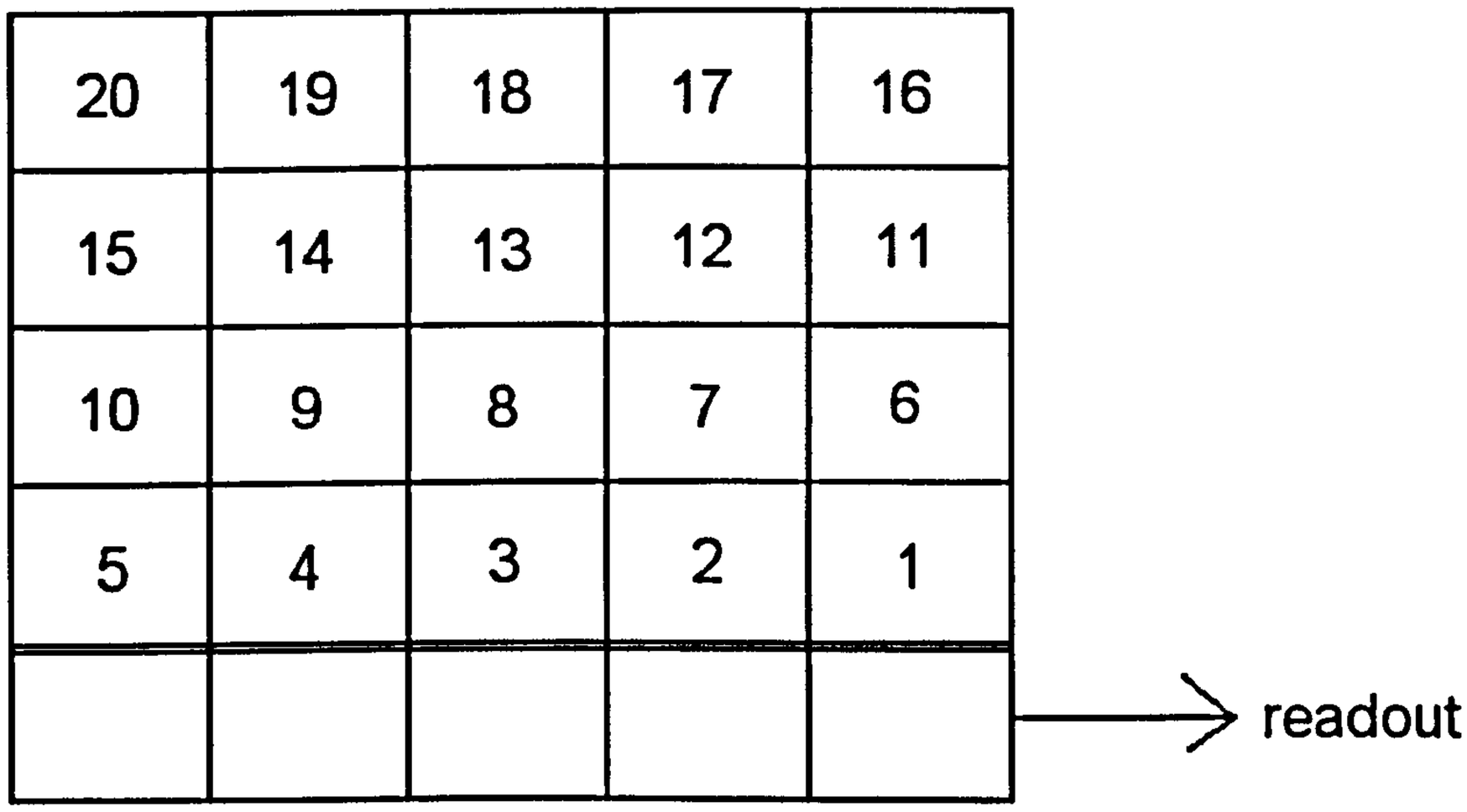

Figure 19 – the order in which pixels are read out from a CCD array. The array shown is very low-resolution for simplicity.

Figure 19 illustrates how the charge is read out from the array. Each row shifts downwards, then the pixels shift horizontally along the readout row until the row is emptied. The process repeats until the whole image is read out.

2.1.4 Saturation

A CCD pixel can only hold a certain amount of charge, called the full-well capacity, before extra charge spills over into the surrounding sensor, rather than being held within the pixel's potential well. The parts of the range from no charge to full-well charge are not equally useful, because at some fraction of maximum capacity the response of the CCD becomes nonlinear (as shown in Figure 20 below), and the response to incident light decreases. This can reduce contrast in highly-exposed areas of an image, and is termed saturation in the sensor.

Image removed due to third party copyright

Figure 20 - graph to show CCD nonlinearity[44]

A second type of saturation can occur in the conversion to digital output if the gain is high enough that the bit-depth of the output is insufficient to represent the pixels' dynamic range.

Saturation is a problem for stereoscopic imaging in cases where saturation sets in at different light levels between the cameras. If one camera saturates before another, different colours will be produced in the two cameras for the same object, and this requires a complex correction during post-processing to remedy. Areas of saturation can also be difficult to match stereoscopically, due to the lack of contrasting features within them.

Care should be taken to avoid saturation (in either the pixels or the control electronics) by choosing appropriate exposure times and gain settings.

2.1.5 Quantum noise (also known as shot noise)

This is the noise inherent to any signal which is made up of discrete quanta, in this case photons[39]. The standard deviation of the number of quanta received is derived from Poisson statistics as the square root of the mean number of counts. This noise implies a signal to noise ratio of the square root of the mean number of counts, and so the relative amount of noise decreases with an increasing number of photons collected. In any imaging device, this means that the deeper the well (the more photons that can be counted by a pixel before saturation), the better the signal to noise ratio can be.

A high signal to noise ratio is usually desirable in images. Specifically, high SNRs provide a more accurate representation of the scene captured, and so exposures should be long enough that as many photons are captured as possible before effects such as saturation arise, in order to minimise image noise.

2.1.6 Charge Transfer Efficiency

CCDs read the images out by means of charge coupling, but the interface between the silicon substrate and the insulating oxide layer is imperfect, which means that some fraction of electrons are lost each time they shift across. The fraction of electrons that make it across each time is called the Charge Transfer Efficiency (CTE), and should ideally be as close to one as possible. CTE is often expressed using Charge Transfer Inefficiency (CTI), which is 1-CTE, since the CTI is typically of the order of 10^{-5}.

If the CTE is less than one, not only will the count of each pixel be reduced, but also the charge that gets left behind will add to the charge that moves into that cell. This is significant in high contrast situations where a very bright pixel is transferred out of a cell, and a dim pixel is transferred in. The electrons left behind by the former can significantly add to the count in the latter, and this shows in images as a "tail" behind bright pixels. This is most problematic in sensors where the well depth is large, so that the inefficiency translates into a number of electrons which is significant compared to the counts in dark areas. The number of charge transfers necessary to read a pixel out will determine the fraction of its charge it is likely to lose.

Dim tails behind bright pixels in dark areas are unlikely to be problematic in normal stereoscopic images, because the effect is usually so small, but may produce slight colour problems in

certain types of colour-masked sensor. Extremely bright pixels in areas of almost total darkness rarely occur in well-illuminated images (unlike, for example, in astrophotography), but may occur on a colour masked sensor where the scene contains a colour which doesn't transmit through one of the colour filters. Some typical R, G and B filter transmissions are shown in Figure 21.

Figure 21 - typical spectral response of R, G and B filters used for colour masking. Figure from Sony[40].

For example, at around 615nm, there is almost zero transmission in the B filter, but over 85% transmission in the R filter. If an object emitting at this wavelength were imaged by a sensor where R and B pixels were adjacent, charge transfer inefficiency could lead to incorrectly high values of the affected B pixels. However, since the G filter has a transmission of greater than 4% for all visible wavelengths, if R and B pixels do not occur in the same lines, then this effect will not occur, such as in Bayer-patterned sensors[41].

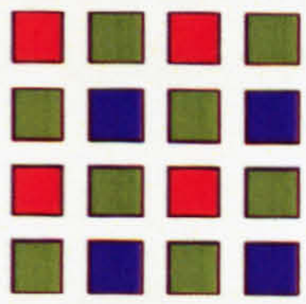

Figure 22 - the Bayer colour mask pattern

2.1.7 Bad pixels

Bad pixels in a sensor array can cause some of the most serious image defects. An array may contain some pixels that have an unusually low response, or an unusually high background count. These not only distort the signal for that pixel, but also often have a very poor CTE and so affect the whole line of pixels behind them, producing lines across images taken with the device. Physically this can only be remedied by replacing the CCD, though image processing software can attempt to disguise the defects.

Bad pixels do not have matching counterparts in the other camera of a stereo pair, and so produce image pixels which have no stereo correspondence. These would be seen as artefacts in the stereoscopic image, which may cause problems during stereoscopic viewing.

2.1.8 Quantum Efficiency (QE)

This is defined as the ratio of photons incident on the detector to photons detected. It would ideally be as close to unity as possible, but is lowered by reflection and absorption in non-sensitive regions of the detector, such as the electrode or below the potential well. This can be improved at long wavelengths by using more transparent electrodes (e.g. polysilicon), or more recently at shorter wavelengths as well by using indium tin oxide gates[42]. The response with wavelength is non-uniform, and the relationship for one type of CCD is shown later in Figure 56.

Another way around this problem is to illuminate the CCD from the back instead of the front – while this does improve QE, it requires the device to be considerably thinned. This adds complexity to the manufacturing process, reducing yield and increasing cost[43], hence it isn't ideal for a consumer product.

A high quantum efficiency is important for stereoscopic imaging, because it reduces the necessary exposure time for an image. If exposure times can be minimised, so can the apparent motions of moving objects in a scene during the exposure, so that motion blur is less apparent.

2.1.9 Blooming

If more electrons than the full-well capacity are generated in a pixel, they spill out into the surrounding pixels, appearing as continuous saturated sections of columns in an image. Blooming can be avoided by using a short enough exposure time that excessive charge does not accumulate, or by use of anti-blooming chip designs. Anti-blooming gates drain away excess amounts of charge so that it does not spill into neighbouring pixels, though this comes at the expense of a reduction in active area (and therefore QE) of around 30%[44].

Blooming constitutes a problem in stereoscopic images where it occurs. Blooming features, apart from being artificial and unrepresented in normal human vision, are unlikely to match as precisely as the normal parts of the images.

Use of anti-blooming sensors allows scenes with high contrast to be captured without underexposure in dark areas or blooming in bright ones. However, any pixels which have made use of the anti-blooming gates have saturated, so measures to deal with saturation (as outlined in 2.1.4) should, where possible, be employed in any situation where blooming may potentially occur. If saturation can be eliminated, the more serious artefacts caused by blooming will also be prevented.

2.1.10 Pixel response

The pixels in the array generally do not all have exactly the same QE. This means that if the chip were to be uniformly brightly illuminated (to minimise the effects of quantum noise, as discussed in 2.1.5), the image produced would not be uniform, and would show a noisy or blotchy effect. This can be substantially compensated for in post-processing using a flatfield – an image taken at uniform bright illumination (to minimise quantum noise) which is used to normalise each pixel's response. Multiple images are sometimes used for flatfields, to further reduce quantum noise levels.

If pixels in the sensors of a stereoscopic camera have a widely varying response, features with a size of one pixel may appear in one image but not another. This may present problems for stereo viewing because may not be possible to stereoscopically interpret a feature that does not appear in both images. If pixel response is sufficiently non-uniform, flatfielding may be necessary.

2.1.11 Dark current

This is the response of the CCD due to electron liberation through thermal excitation, and means that over the duration of the exposure there will be a cumulative background count dependent on the temperature and the exposure time. This can be reduced by cooling the chip, but since this project is considering exposures of a fraction of a second with reasonably bright scenes (as opposed to, for example, astrophotography), the noise levels at room temperature do not cause significant problems[45].

Dark current can reduce the dynamic range, and therefore contrast, of the camera output if it is of a high enough magnitude, which makes features harder to distinguish, and may need special attention if images are being taken of dark scenes with long exposure times.

2.2 Single camera calibration

Camera calibration has been a field of research for over 100 years. As techniques have become more sensitive, the parameters of camera systems have been measured more accurately, and different types of distortions introduced by lenses have been uncovered[46].

2.2.1 Understanding the terminology behind image formation

There are two approaches and sets of possible terminology to use when discussing image formation by lenses and sensors.

The first is a treatment where the sensor determines the main axis (called the principal axis) of the system as the normal to the surface of the sensor which passes through the theoretical pinhole This is a construction that represents the centre of an ideal lens system, through which light rays pass undeviated. Light from the scene enters through the pinhole, from which the rays emerge to form the image. The point at which the principal axis intersects the image plane is called the principal point, and the distance from the principal point to the pinhole is called the principal distance. Distortion effects due to the lens are added afterwards, and if the distortions are centred around a different point to the principal point, a linear offset is applied to allow the distortions to be modelled using the principal point as their centre.

The second is a treatment where the lens determines the main axis (called the optical axis) of the system as the axis along which distortions due to the lens are centred. Instead of treating the lens as a pinhole, the lens instead is modelled as having two nodal points, one front and one rear. Light from the scene enters through the front nodal point, and emerges from the rear nodal point to form the image. The point at which the optical axis intersects the image plane is called the optical centre (also sometimes confusingly called the principal point), and the distance from the optical centre to the rear nodal point is called the effective focal length. All lens-based aberrations are modelled using the optical centre as their centre.

Choosing which approach to use depends largely on the relative importance of lens and sensor effects. The principal axis approach has the advantage of ease of measurement, because the sensor is an easier object to physically measure than the nodal points of a lens. The optical axis approach is a truer reflection of reality, because it does not need to make offsets or approximations in the distortion centre to fit the chosen axis. However, because the optical axis relies on distortions for its definition, it is difficult to determine accurately in cases where the distortions are small. Because low distortion lenses have specifically been chosen in section 1.7, and are preferred for use in a stereoscopic camera because they cause smaller false disparities, it is more appropriate to use the principal axis approach in describing how images are formed by such cameras. Where principal point is used in this thesis relation to lens aberrations such as distortion, a linear offset as described here from the position defined in section 2.2.2 is implicit.

A more detailed description of the principal axis and principal point concepts follows in sections 2.2.2 and 2.2.3.

2.2.2 Principal point

The principal axis of a camera is the perpendicular to the image plane which passes through the lens centre (theoretical pinhole or in practice, the rear nodal point). The principal point is the point at which the image plane and the principal axis intersect. The distance from the principal point to the lens centre is called the principal distance, which is approximately equal to the lens focal length at infinity focus. Figure 23 illustrates the geometry:

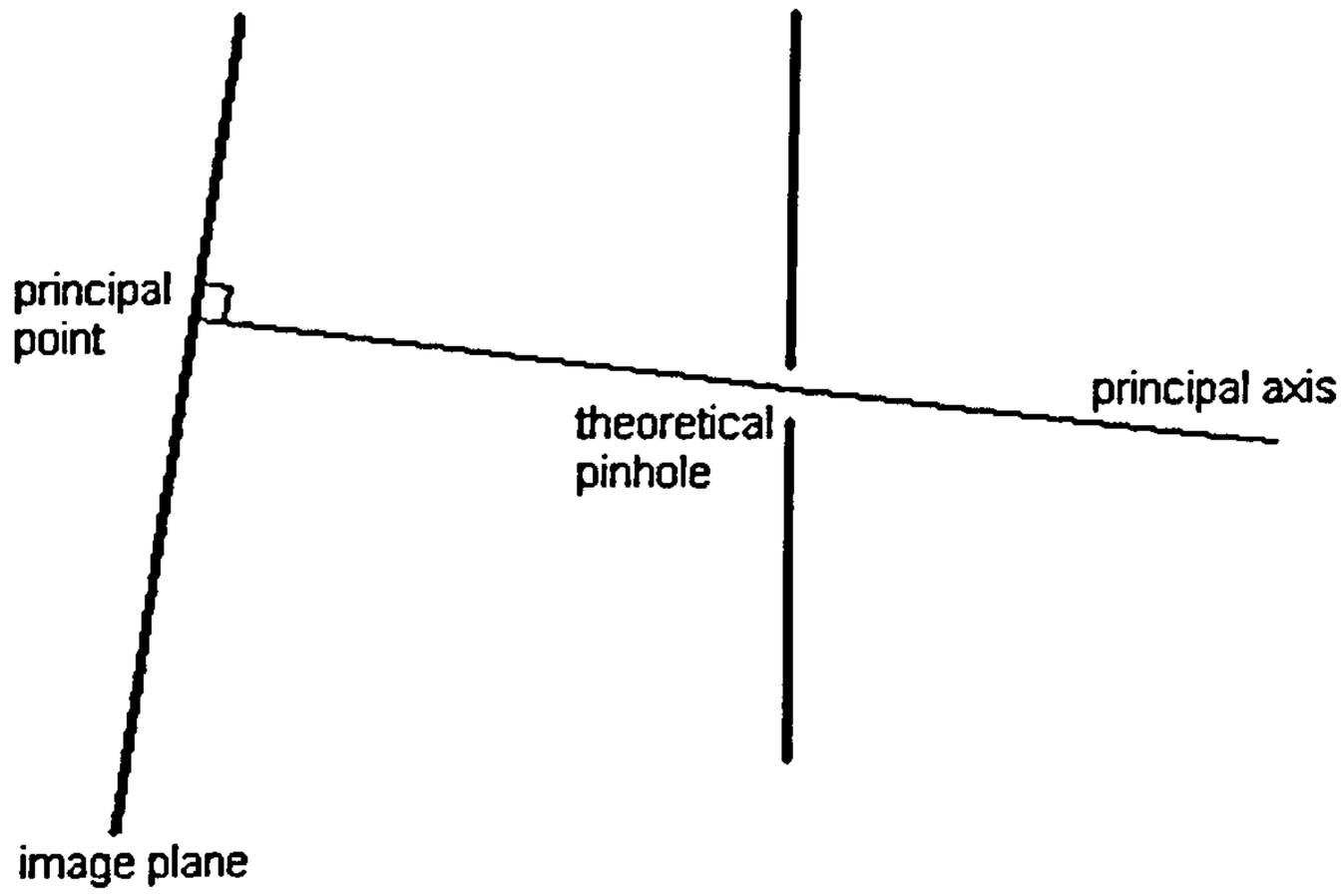

Figure 23 – principal axis and principal point

The principal point is difficult to find because it is defined by the lens centre, which is hard to locate physically[47]. Many methods have been suggested for finding it, with varying degrees of accuracy. Photogrammetric methods used for aerial cameras can locate the principal point to around 10μm[47], but these are not readily applicable to normal machine vision systems, which are not built to the same standards. The methods traditionally used for most cameras are much less accurate, because the construction of machine vision systems is typically inferior to photogrammetric systems in terms of rigidity, resolution, component quality and initial component positioning. These deficiencies would outweigh the advantages of an accurate calibration procedure.

It is often important to know the location of the principal point in stereo photography because images need to be matched to each other, and the principal point is one of the parameters which determines the transformation from object space (the three-dimensional space of the physical world) to image space (the two-dimensional space containing the image plane). If the principal points between stereo cameras differ by an unknown amount, the alignment of those images requires feature recognition, which can be imprecise and unreliable.

2.2.3 Principal distance

The principal distance of a camera effectively determines the scale of an image. Increasing this distance decreases the amount of the world that the imaging surface is exposed to, zooming the image in. The principal distance is another parameter that determines the mapping of object

space to image space, and in a stereo camera the principal distances (or at least their ratio) need to be known in order to rectify the images for display.

2.2.4 Lens distortion

The main component of lens distortion is usually radial, and can be defined relative to the principal point of the camera[26], the radius being a line between the image point (the location of an imaged object point in image space) and the principal point. This distortion moves an image point along its radius by an amount dependent on the radial distance, and is illustrated in Figure 24.

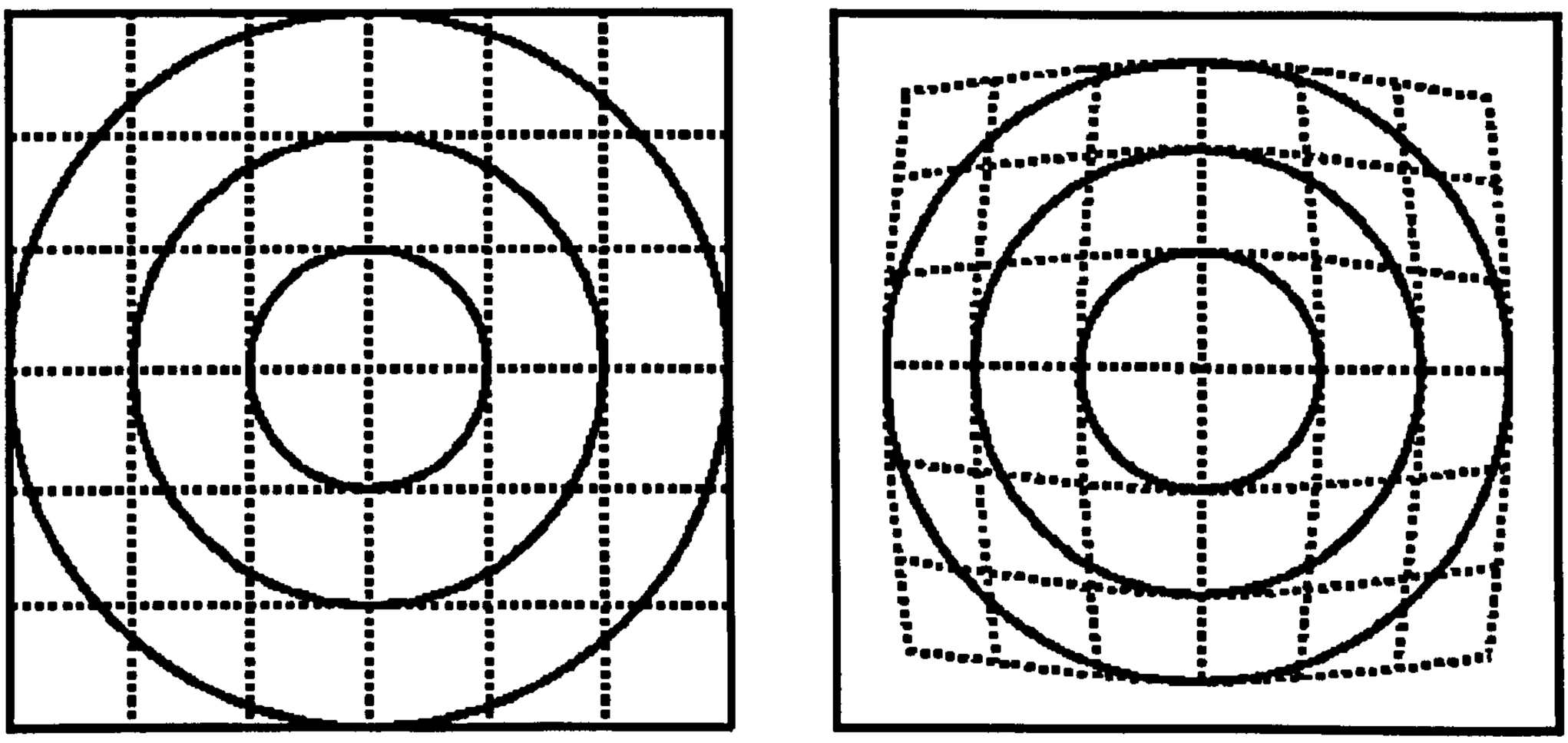

Figure 24 – Image before and after radial distortion

In early calibration work, this distortion was partially balanced out by varying the assumed value of principal distance[47], because a radial distortion has an overall zooming characteristic. But this method was flawed because the zoom caused by radial distortion varies with radius, unlike that caused by varying the principal distance value.

Radial distortion is now modelled as a polynomial function[26]:

$$\Delta r = r.(k_1 r^2 + k_2 r^4 + k_3 r^6 + \ldots)$$

Equation 1

where Δr is the radial distortion, r is the image point radius, and the k_n are the radial distortion coefficients. Often the distortion can be modelled using just the k_1 coefficient, but the k_2 and k_3 coefficients are sometimes used for better accuracy or with wide-angle lenses.

Lens distortion also has a tangential component, sometimes called decentring distortion because it can be caused by decentring of lens elements. The effect is similar (though not identical) to adding a small prism to the lens system[47]. Tangential distortion moves an image point tangentially by an amount dependent on both its radius and its radial angle.

Tangential distortion can be modelled as[26]:

$$\Delta x = P_1.(y_r^{\,2} + 3x_r^{\,2}) + 2P_2.x_r.y_r$$
$$\Delta y = P_2.(x_r^{\,2} + 3y_r^{\,2}) + 2P_1.x_r.y_r$$

Equation 2

where x_r and y_r are Cartesian coordinates relative to the principal point, and P_1 and P_2 are the tangential distortion coefficients. The profile function P(r), which is the tangential distortion at a radial distance r (equal to $\sqrt{(x_r^2+y_r^2)}$), along the axis of maximum tangential distortion, is related to the coefficients by:

$$P(r) = r^2\sqrt{P_1^{\,2} + P_2^{\,2}}$$

Equation 3

Lens distortion is often a necessary effect to account for in an accurate camera calibration[48], and can be compensated for to varying extents by using a least squares fit with the above models, or by other novel methods[49]. An offset to the principal point can also be used to compensate for tangential distortion[26], though this is not an exact correction.

2.2.5 Camera parameters

The intrinsic (relating to the camera's interior) and extrinsic (relating to the camera's exterior) parameters are directly related to the view of the object space. To derive these parameters, it is sufficient to observe a scene with a number of features whose positions relative to the camera are known. Standard calibration targets can be used for this purpose, as can natural scenes on Earth that have been carefully surveyed, and even stellar images have been used for calibration. A simple pinhole camera model can be used as shown in Figure 25 below:

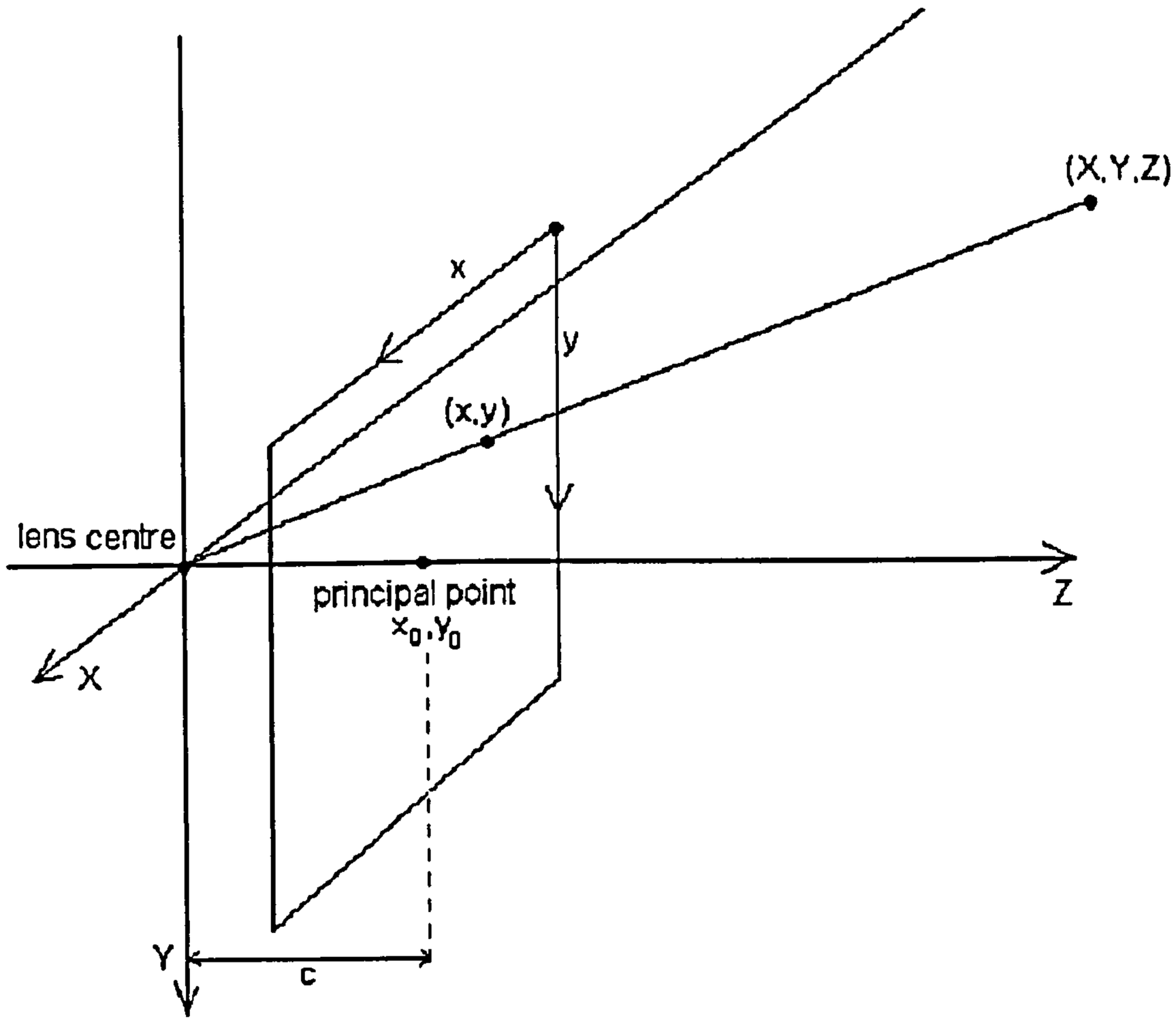

Figure 25 - pinhole optical geometry

Figure 25 shows a point in object space with object (or world) coordinates (X,Y,Z), the lens centre defined as the origin of the X,Y,Z object coordinates, and a front-projected image plane with image coordinates x,y, at a Z-distance c from the lens centre. The origin of the image coordinates is at the top left of the image, and the object point produces an image point in the image plane, at (x,y).

The front-projected image plane is a construction used to simplify the description of the imaging device. The actual image is formed in a plane behind the lens centre (with at -c on the Z-axis) and is inverted. Reflecting the image plane in the Z=0 plane gives the front-projected image plane, in which the X and x directions and the Y and y directions are aligned.

The relationship between the object space and image space can be expressed as:

$$x = x_0 + c.p_x.\frac{X}{Z}$$

$$y = y_0 + c.p_y.\frac{Y}{Z}$$

Equation 4

where x and y are the image coordinates (in pixels, in the detector plane) relative to the principal point, x_0, y_0, c is the principal distance, p_x and p_y are the number of pixels per unit length, and X, Y and Z are world coordinates defined relative to the image plane, with the origin at the lens centre.

2.2.6 Camera calibration

Least squares or other optimisation methods can be used to derive the camera parameters given a set of known points in space corresponding to measured features in an image[58]. This is known as camera calibration, and relies upon the use of a calibrated target to define these control points[57]. The accuracy with which the spatial positions of these control points are known partly determines the accuracy of the camera calibration.

Camera calibration can take one of two approaches. The target may be three-dimensional[50], in which case one image may be sufficient for calibration. Alternatively, the target may be two-dimensional[110], in which case multiple views are required.

The requirement that a calibrated target must be used is generally restrictive[57], especially in the case of three-dimensional targets, although the results can be highly accurate[110]. As with all calibration schemes where a target is used, the calibration may not perform well outside the spatial range defined by that target[74]. As described in 3.2.1, calibration is generally susceptible to errors introduced by coupled parameters. Some methods derive the rotational and translational parameters separately using measurements such as vanishing points of parallel lines[51], which do not depend on position. The accuracy of such methods is so far not as good as for conventional methods[51,110].

2.2.7 Self-calibration

Also known as autocalibration or bundle adjustment, this technique uses multiple images of a target to determine both the camera parameters and the world coordinates of the target points[52]. The 10-parameter model used for digital camera self-calibration is[53]:

$$x_r - \frac{x_r}{c}.\Delta c + x_r.r^2.k_1 + x_r.r^4.k_2 + x_r.r^6.k_3 + (2x_r^2 + r^2).P_1 + 2P_2.x_r.y_r + b_1.x_r + b_2.y_r = -c.\frac{X}{Z}$$

$$y_r - \frac{y_r}{c}.\Delta c + y_r.r^2.k_1 + y_r.r^4.k_2 + y_r.r^6.k_3 + (2y_r^2 + r^2).P_1 + 2P_2.y_r.x_r = -c.\frac{Y}{Z}$$

Equation 5

Δc is the change in principal distance (expressed as a decimal fraction), b_1 and b_2 are empirical correction terms for differential scaling and non-orthogonality between x and y axes, and other symbols are as defined above. This can be seen to be a modification of the simple pinhole model in the section above, to account for lens distortion, changes in principal point and distance, and non-orthogonality, and is not specific to self-calibration techniques. The two b terms can be omitted for digital CCD cameras with square pixels to leave an 8-parameter model, and for low-accuracy applications, only the 4 parameters of principal point, principal distance, and first radial distortion coefficient[26] are used.

Self-calibration finds the parameters in the above equation without knowing the world coordinates X, Y and Z, which is useful in situations where the target itself cannot be measured accurately[54]. All that is known about the target is that it remains the same between images. There are different ways to apply this technique and solve for the camera parameters, such as using multiple views of flat targets[55] or using camera motion confined to a plane[56] - usually the method will use some kind of constraint to aid in an accurate calibration. One of the conclusions from self-calibration is that the measurement of image points is more important for camera calibration than having a well-calibrated target[57].

Self-calibration has many advantages over conventional calibration, but care must be taken with the results. If measurements are constrained in some way not accounted for in the method, one or more parameters may be poorly defined[58]. Though the technique can be very versatile and does not require a calibrated target, this is often at the expense of accuracy in determining the calibrated parameters[59].

2.2.8 Types of self-calibration

Self-calibration can be performed in a variety of ways, but can be broadly classified into three types.

The first is global nonlinear minimisation, and covers calibrations where values for all camera parameters are found by an iterative minimisation algorithm. The quantity to be minimised varies with the calibration scheme, but is generally a measure of the errors in the equations relating the scene and image, for a given set of camera parameters. The nature of this process means that initial estimates for the parameters are required, and if these are not close to the actual values, the algorithm may converge to a local rather than a global minimum, giving incorrect results[59]. Applications of this type of calibration are numerous[52,60,61,55,57], and where the accuracy is quoted[55,57], the accuracy is generally good.

The second method is the closed form solution, where the calibration is used to determine all the camera parameters through solving linear equations. Such algorithms are not as computationally intensive as nonlinear minimisation methods, and the calibrated parameters are not affected by an inaccurate initial estimate. The main drawback is that noise in the extraction of calibration features leads to inaccuracy in the calibration parameters[59]. Applications using this type of calibration[62,109] are less numerous than those using minimisation techniques, and where a method's accuracy is quoted[109] it tends to be relatively poor.

The third type is a mixture of the first two, known as a two-step solution. Some parameters are found by analytical methods, while others are calculated using a nonlinear minimisation algorithm. Parameters found by analytical methods are usually determined first, because they help to constrain the parameters found by the optimisation. Applications of this type of calibration are numerous[63,101,64,74,85], and where accuracy is quoted[74,85] the results are generally quite good.

Because self-calibration does not require a calibrated target, it is relatively easy to perform. However, the above variety of methods illustrates how a balance must be struck between accuracy and stability of any self-calibration scheme.

2.3 Stereoscopic camera design

2.3.1 Basis for classification of stereoscopic cameras[25]

Stereoscopic cameras exist in a variety of forms, but they fall into four simple categories based upon the number of lenses and sensors employed. Cameras can have either one sensor or multiple sensors, and one lens or multiple lenses. The merits of these designs are discussed below.

2.3.2 Single lens, single sensor[25]

This is perhaps the simplest stereoscopic camera design, and consists of one lens and one sensor combined in a single camera unit, which takes two images of an object from different viewpoints. Figure 26 shows an example of such an arrangement.

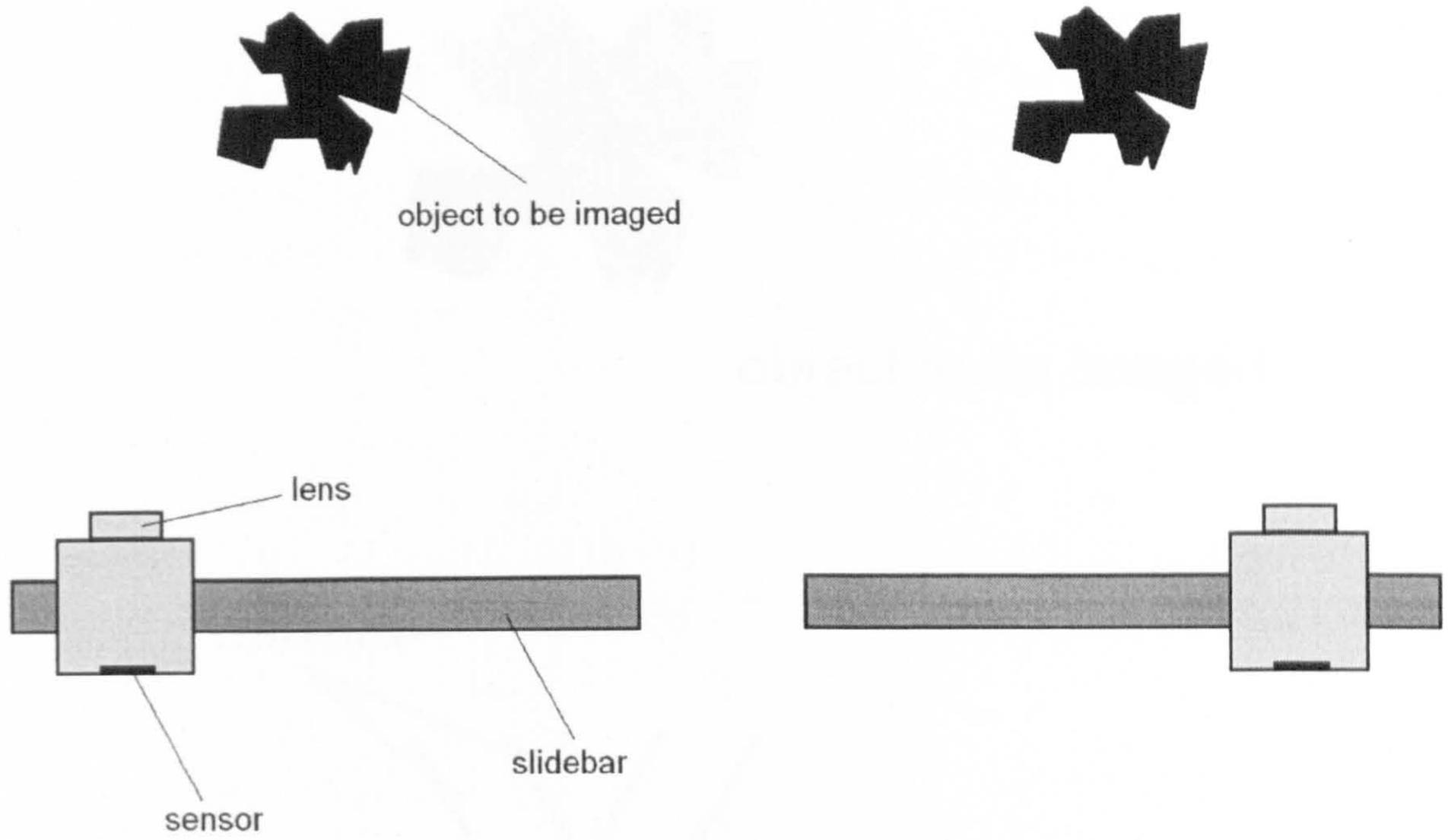

Figure 26 - single lens, single sensor stereoscopic camera. The camera takes two images from different horizontal positions

Arrangements based on this principle[65, 9], are in many ways ideal for stereoscopy. In terms of parameters such as rotation, translation, principal point and principal distance the two views are identical except for the horizontal translation between the views. Calibration must only address the issue of an accurate translation mechanism, and depending on the system perhaps a lens calibration.

However, there is one significant disadvantage to this method of acquiring stereoscopic images. The two views are different in one important respect, the time at which the view is captured. This makes this arrangement highly unsuitable for imaging any scene for which there is significant motion between the views being captured, as this will introduce significant stereoscopic errors.

2.3.3 Single lens, multiple sensors[25]

This design uses a single lens to form two simultaneous images. The two viewpoints are usually created by using lenses or mirrors to split the field of view of the main lens. An example of this is shown in Figure 27 below.

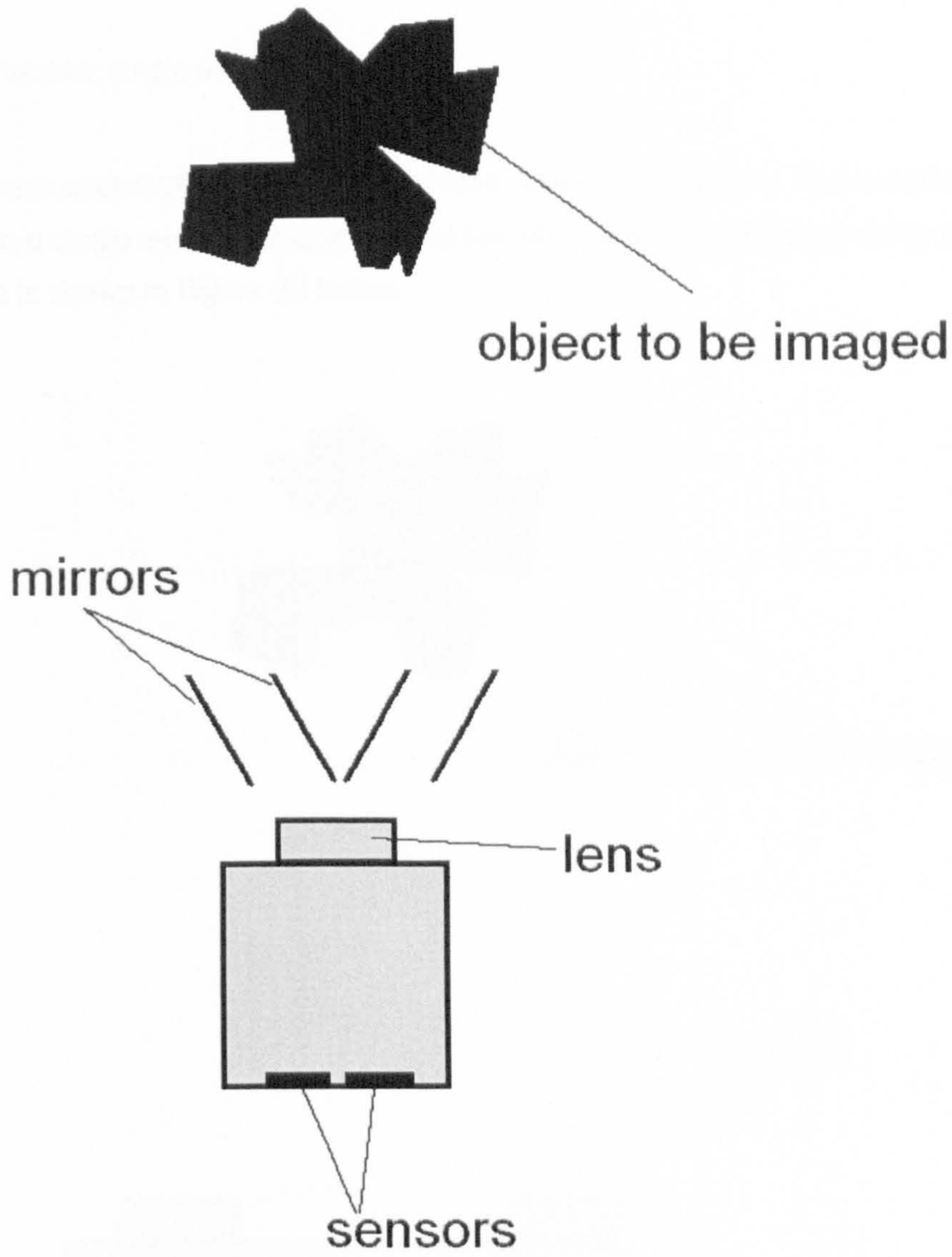

Figure 27 - single lens, multiple sensor stereoscopic camera. The mirror arrangement creates two viewpoints which are imaged on to the two sensors through a single lens

Arrangements of this type[66] are popular with amateur enthusiasts, because mirror assemblies can often be attached to ordinary cameras at low cost. Under such circumstances, a single detector (often film) acquires two different images side by side.

Multiple components need to be aligned in the mirror or secondary lens arrangement, and the primary lens must produce acceptable images far away from the optical axis. There are no particular synchronisation problems, but crosstalk (leaking of image information between the two views) is a possible drawback.

2.3.4 Multiple lenses, single sensor[25]

This arrangement uses multiple lenses which have different viewpoints. The images are placed side by side on a single sensor using an optical folding system. An example of this type of stereo camera is shown in Figure 28 below.

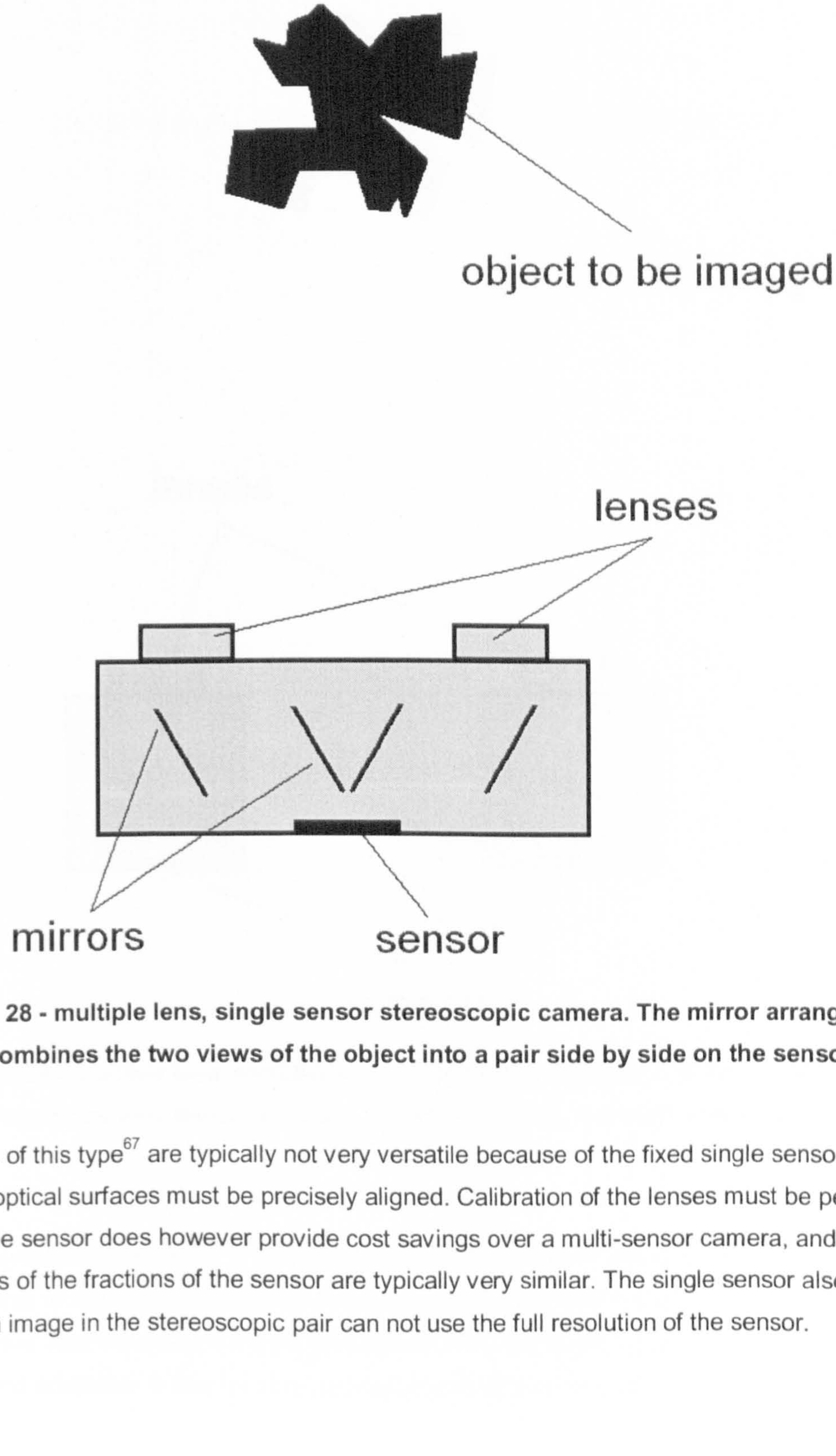

Figure 28 - multiple lens, single sensor stereoscopic camera. The mirror arrangement combines the two views of the object into a pair side by side on the sensor

Cameras of this type[67] are typically not very versatile because of the fixed single sensor, and multiple optical surfaces must be precisely aligned. Calibration of the lenses must be performed. The single sensor does however provide cost savings over a multi-sensor camera, and the properties of the fractions of the sensor are typically very similar. The single sensor also means that each image in the stereoscopic pair can not use the full resolution of the sensor.

2.3.5 Multiple lenses, multiple sensors

This type of stereoscopic camera uses multiple lenses, each with their own image sensor. The lens-sensor systems are typically contained inside individual camera bodies. An example of this type of system is shown in Figure 29 below.

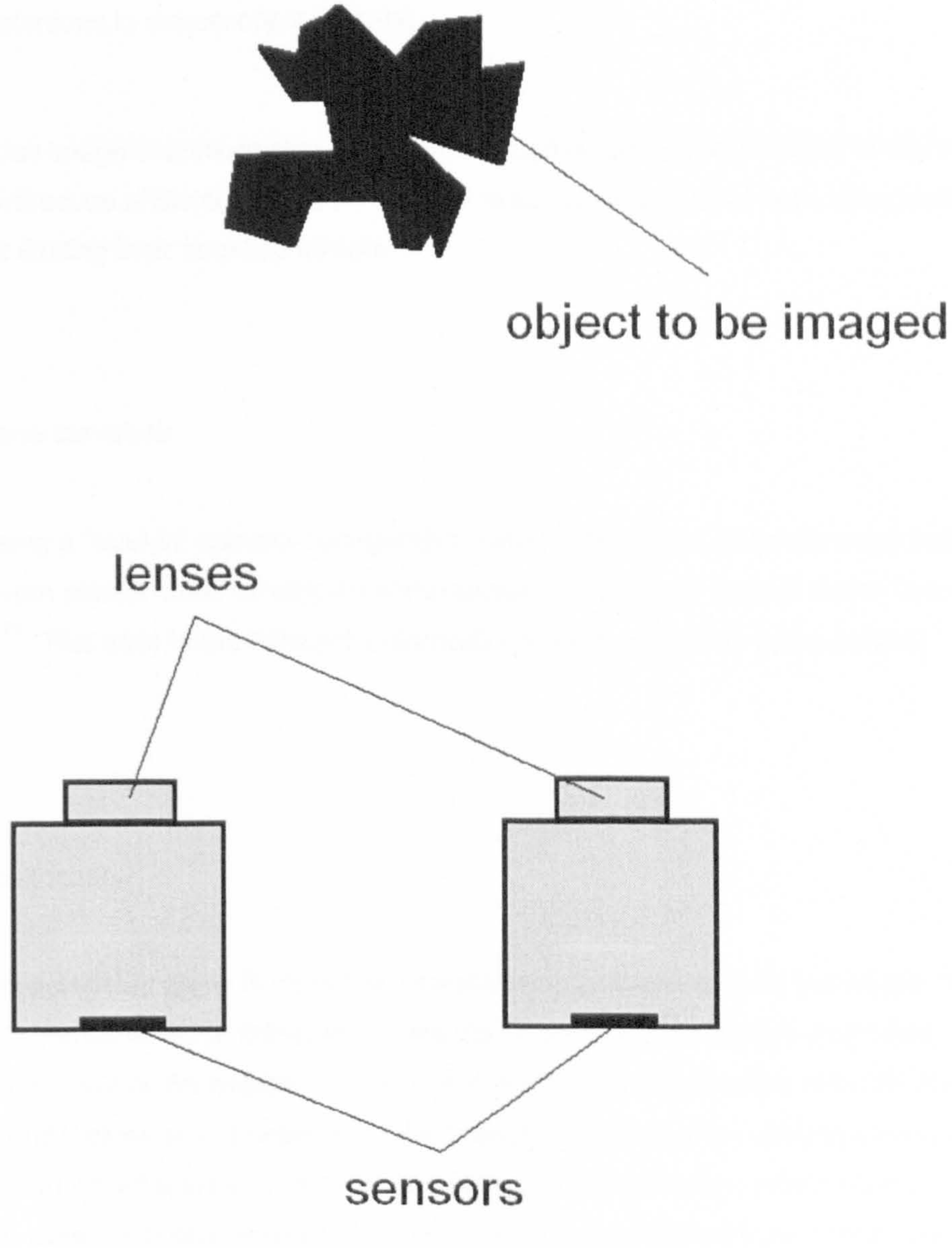

Figure 29 - multiple lens, multiple sensor stereoscopic camera. Each camera head, consisting of one lens and one sensor, is self contained, and each views the scene from a different viewpoint

Camera systems of this type[68,74] tend to have the highest component cost, and have multiple components to align. Lenses must be matched or calibrated, and the sensors must be aligned or calibrated also. However, the arrangement produces full sensor resolution images, and is flexible and adaptable in that camera separation and relative orientation can be varied without

changing the optical paths enclosed in the instrument. Such a system can also be easily assembled from single camera components.

2.4 Stereoscopic camera calibration

2.4.1 Image distortions in stereoscopic imaging

There are various image distortions that can arise during the capture and display of stereo images which introduce artificial disparities[69,70]. The major ones are listed here, along with some suggestions for limiting their negative effects.

2.4.2 Depth plane curvature

Arises when using a "toed-in" camera configuration (where the optical axes converge before infinity). The depth planes curve so objects at the centre of the image appear closer than ones at the edges[69,70]. This error in the horizontal disparity can be removed by using parallel cameras.

2.4.3 Depth nonlinearity

Occurs if the object infinity plane is displayed anywhere other than infinity in the image. This brings the depth planes forward, compressing the depth between the display and infinity, and stretching depth in front of the display[69,70]. Though this isn't a major problem with still images, it may look odd, and it gives an impression of false velocity changes when viewing stereoscopic video. The solution would appear to be to display object infinity at image infinity (one of the properties of so-called orthostereoscopic display[23]), but this conflicts with the human factors requirement that the depth be constrained to a volume near the display itself[70].

2.4.4 Shear distortion

This arises when the observer moves around while viewing a stereo image. If he moves laterally, images in front of the display plane appear to follow him, while images behind it move

in the opposite direction[69]. This produces an apparent motion of a still image, and the only obvious solution is for the observer not to move. This distortion has no effect on stereo fusion, but is an unnatural effect to observe during viewing.

2.4.5 Depth magnification

Errors occur when the image is enlarged or reduced for presentation. Since depth is not linearly related to the x and y dimensions of the image, a magnification in the x and y dimensions does not produce a correct magnification in depth[69]. This can change the amount of perceived depth when images are viewed at different sizes. The only proper solution is to tailor the image capture to the display, though the development algorithms to vary depth independently of x and y would be another approach.

2.4.6 Keystone distortion

This is another distortion that is caused by using a toed-in (ie non-parallel) camera configuration. If the camera axes are non-parallel, the image planes are also not parallel, and so the sides of each camera's image planes appear at different distances from the observer[69,70]. This introduces a significant amount of false vertical disparity at the edges, and the solution is to use parallel cameras.

2.4.7 Lens distortion

This is also known as barrel and pincushion distortion, and is one of the Seidel aberrations[71,72] (and most significant one for multiple lens assemblies, since the other aberrations can be largely corrected[89]). It may arise from using spherical lens elements, and has a large effect on the corners of images, either stretching them outward (pincushion) or squashing them inward (barrel). This gives rise to both horizontal and vertical disparities, and is a big problem with some short focal length lenses. The solution is to use better lenses, sometimes helped by aspherical elements[71] in the lens assembly. Lens distortion is described above in section 2.2.4.

2.4.8 Brightness and colour variation

These cause problems in recognising point correspondences, especially in areas of gradual brightness or colour gradient in images, and can arise under two circumstances. Firstly, one or both views may suffer from vignetting, imposing a brightness change on the images which is not

identical for points in the scene. This can in principle be calibrated out, but it is preferable to use systems with as little vignetting as possible. The second circumstance is where the two views are acquired with a different sensitivity in one, two or all three colours. This can be minimised by matching the optical systems and detectors (or using the same optical system and detector for each image), or by calibration and post-processing (such as white balancing).

2.4.9 Epipolar geometry[86]

Stereo calibration is often based on the determination of the camera pair's epipolar geometry, which is a description of how the two cameras are orientated with respect to each other. The epipolar geometry of a camera pair is illustrated in Figure 30:

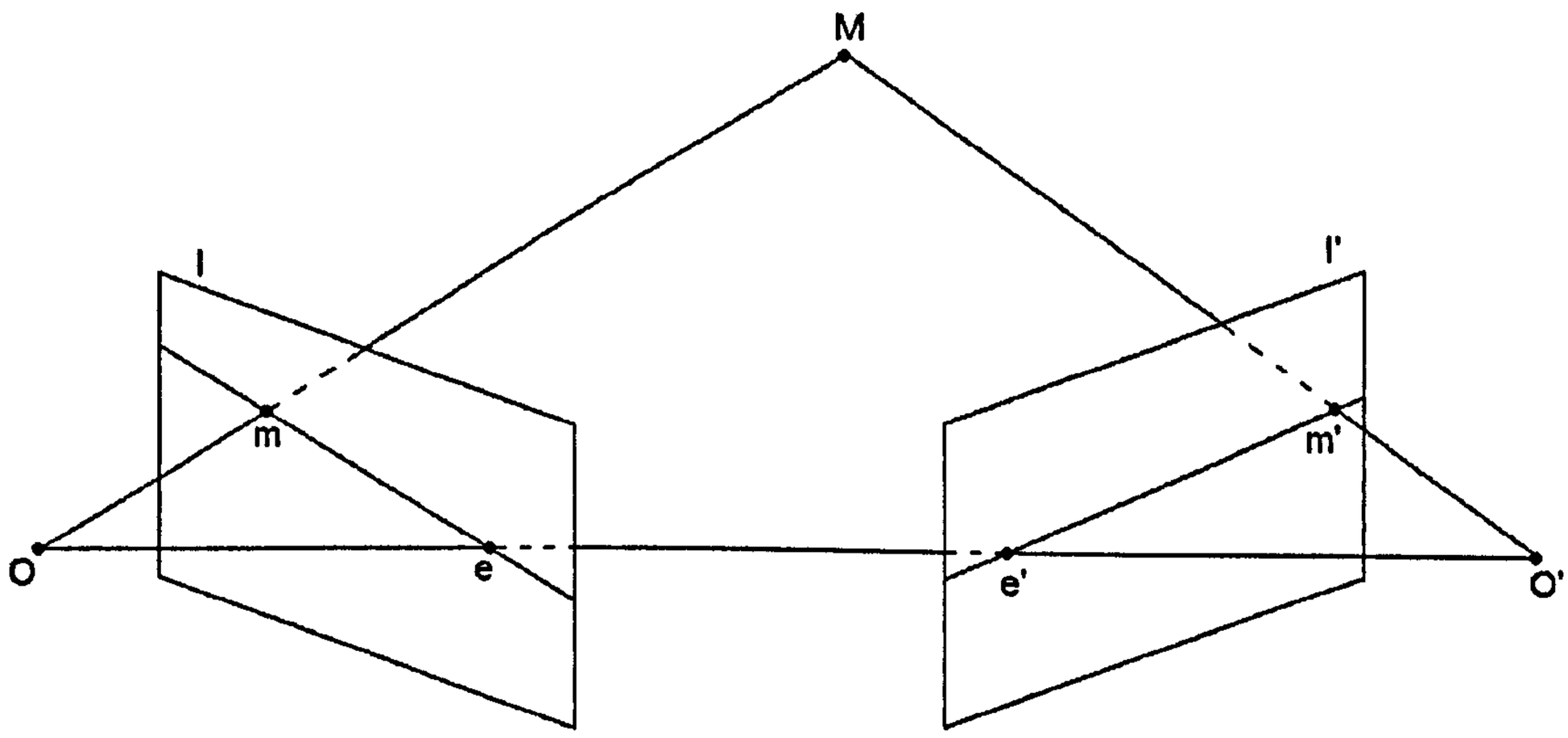

Figure 30 - epipolar geometry

In Figure 30, a point M is imaged by two cameras, with optical centres O and O', and front-projected image planes I and I'. The images of the point M, called corresponding points, are m and m', in pixel coordinates. The points at which the image planes intersect the line OO', marked as e and e', are the epipoles of the system. The lines me and m'e' are called the epipolar lines for points m and m', and the set of planes containing the line OO' are called epipolar planes. The epipoles do not have to lie within the images themselves.

The transformation from I to I' is determined by the fundamental matrix, F. This 3x3 matrix satisfies Equation 6[73]:

$$m'^{T}.F.m = 0$$
$$F.e = 0 = F^{T}.e'$$

Equation 6

The fundamental matrix depends on the properties of the cameras and their relative orientations, is defined up to a scale factor, and has a determinant of zero[74].

2.4.10 Stereoscopic camera calibration

Stereoscopic cameras may be calibrated by simply calibrating each of the two cameras individually[52,55,57,60,61,62,63,64,85,101,109]. However, since some parameters can be defined relative to the other camera in the pair, schemes designed specifically to calibrate stereoscopic cameras can be more efficient.

A typical calibration of a stereoscopic camera depends on recovering a good estimate for the fundamental matrix. The principle behind doing this is to use pairs of known corresponding points (pairs of m and m') to estimate the elements of the fundamental matrix, F[86,83], and from this to deduce the epipolar geometry[75,76]. Schemes using this technique[74,77,78,79,80] are generally less numerous than single camera calibration techniques due to their more specialised nature, but where accuracy is quoted, the calibration results are good[74].

Once the fundamental matrix has been determined, the images can be rectified. Rectification[81] is a warping of the two images such that all epipolar lines are parallel to the x-axis, and that all corresponding points have identical y-coordinates[73]. This step aims to eliminate vertical disparities from the images. A horizontal stretch is also required to ensure that points at the same distance have the same horizontal disparities.

The core problem for stereoscopic calibrations based on estimating the fundamental matrix is that the estimate of the matrix is dependent on the volume of space occupied by the calibration target[74]. This means that if a stereo camera pair's fundamental matrix is estimated from a calibration target in one position relative to the camera, the rectification may be poor for objects imaged in other positions relative to the camera. This may be addressed through performing calibrations with targets at a variety of distances, and using the fundamental matrices obtained to estimate a fundamental matrix that enables a better rectification across a range of imaged volumes[74]. However, these compromise estimates of F may still produce unacceptable errors in rectified images. Rectifying images individually, rather than using one initial calibration, avoids problems with the target volume, but introduces two other problems. Firstly, that the target may not always contain convenient or easy points for automatic stereo matching, and secondly that the amount of computing power requires to estimate F may be nontrivial.

The warping involved in rectification also causes a loss of resolution, as well as a slight blur introduced by the necessary pixel interpolations. Using software to sharpen the image cannot

undo the blur, as the necessary image information does not exist. The only way to achieve full sensor resolution would be to revert to the uninterpolated image, thus undoing the rectification.

2.5 Evaluation of existing techniques in stereoscopic camera design and calibration

2.5.1 Evaluation of existing techniques - camera design

In terms of stereoscopic camera design, the design which offers the most potential is the multiple lens, multiple sensor design. This design allows the camera system to be assembled from two individual cameras, which may either be calibrated independently or together as required. The design allows flexibility in terms of camera separation, while robustly maintaining the separation of the outside world and the camera interiors. It also allows for full sensor resolution to be available both in stereoscopic and a monoscopic mode, and stereoscopic synchronisation is trivial to achieve.

The disadvantages of the multiple lens, multiple sensor design are the cost of two complete cameras, and the alignment (both initial and long-term) of the two camera heads with respect to each other. The former can be offset through the use of off-the-shelf components, while the latter can be overcome by use of suitable calibration and mounting techniques.

While the above details of the construction of a two-camera stereoscopic instrument must be refined, the basic properties of this existing system are adequate for use in this project.

2.5.2 Evaluation of existing techniques - camera calibration

Some current calibration techniques[57,74,110,85] have been shown themselves to be able to calibrate a stereoscopic camera to within the requirements of the human visual system, as discussed in 1.6. However, the techniques have some shortcomings and limitations.

None of the existing techniques addresses the varying lens focus settings necessary to produce sharp images for scenes at different distances, and the effect that varying these settings has on the calibration. All are image-based, and as such the above methods suffer from a degree of susceptibility to coupling of calibration parameters. This reduces the accuracy of any such parameters, and means that the accuracy of the calibration is degraded for scenes outside the calibration target point positions. All of the methods which have been shown to meet the human

factors requirements above rely on nonlinear minimisation techniques, which are not only computationally intensive but also susceptible to settling on local rather than global minima.

The above methods are also based on processing images, which makes a real time hardware calibration a cumbersome process. Though the methods can be (and are in the above examples) applied as a correction to images rather than as a means of building the camera system, this correction requires destructive interpolation of images. Though the above calibrations rely on multiple images of a calibration target from different angles and positions, no details of measures taken to ensure the stability of the camera system under these movements (and beyond, into normal use) are given.

2.5.3 Areas for improvement

The following aspects of stereoscopic camera calibration are identified for improvement.

- The calibration scheme should be such that the calibration remains valid when the lenses are focused
- Any rectification of images should not involve interpolative correction
- Coupling of calibration parameters should be minimised
- The calibration scheme should use only linear techniques
- The calibration scheme should be able to vary and set as many camera parameters as possible in real time
- Mounting techniques must be used to allow for camera parameters to be set as required, and to maintain the calibration parameters during use

These improvements must be carried out with the constraint that such a camera and calibration system produces vertical disparity that is low enough to be tolerated by the human visual system.

The process of attempting to implement these improvements will test the thesis hypothesis, restated below:

"It is possible to design and construct a stereoscopic camera system which produces images for display that satisfies the requirements of the human visual system for comfortable, fused stereoscopic viewing. This can be achieved without any knowledge of the particular images for display, and need not involve any interpolative post-processing of the individual colour images."

The structure of the project to address these improvements is as described in 1.8.

3 Basis for calibration of a stereoscopi c camera

3.1 Introduction

This chapter examines the basis for calibrating a stereoscopic camera in hardware. 3.2 explains the problem of coupling of calibration parameters, and suggests a two stage calibration to solve this problem. 3.3 describes an existing technique for principal point location, and describes a novel adaptation of this technique which enables the sensor to be calibrated. The second part of the calibration, calibrating the lens, can be achieved using conventional image-based techniques. 3.4 identifies the parts of the camera system which contain the greatest potential for instability.

3.2 Separation of rotational and translational errors

3.2.1 Justification

Conventional stereo calibrations are difficult to perform accurately because small rotational and translational errors in the image can look very much alike. For example, pitching the camera down by a small amount produces an effect on any images captured which is very similar to the effect produced by translating the camera downwards by a small amount. In theory this effect can be distinguished by imaging objects at different distances (as described in sections 2.4.10 and 3.2.2), since a translation will produce a relative disparity shift between such objects, whereas a rotation will not.

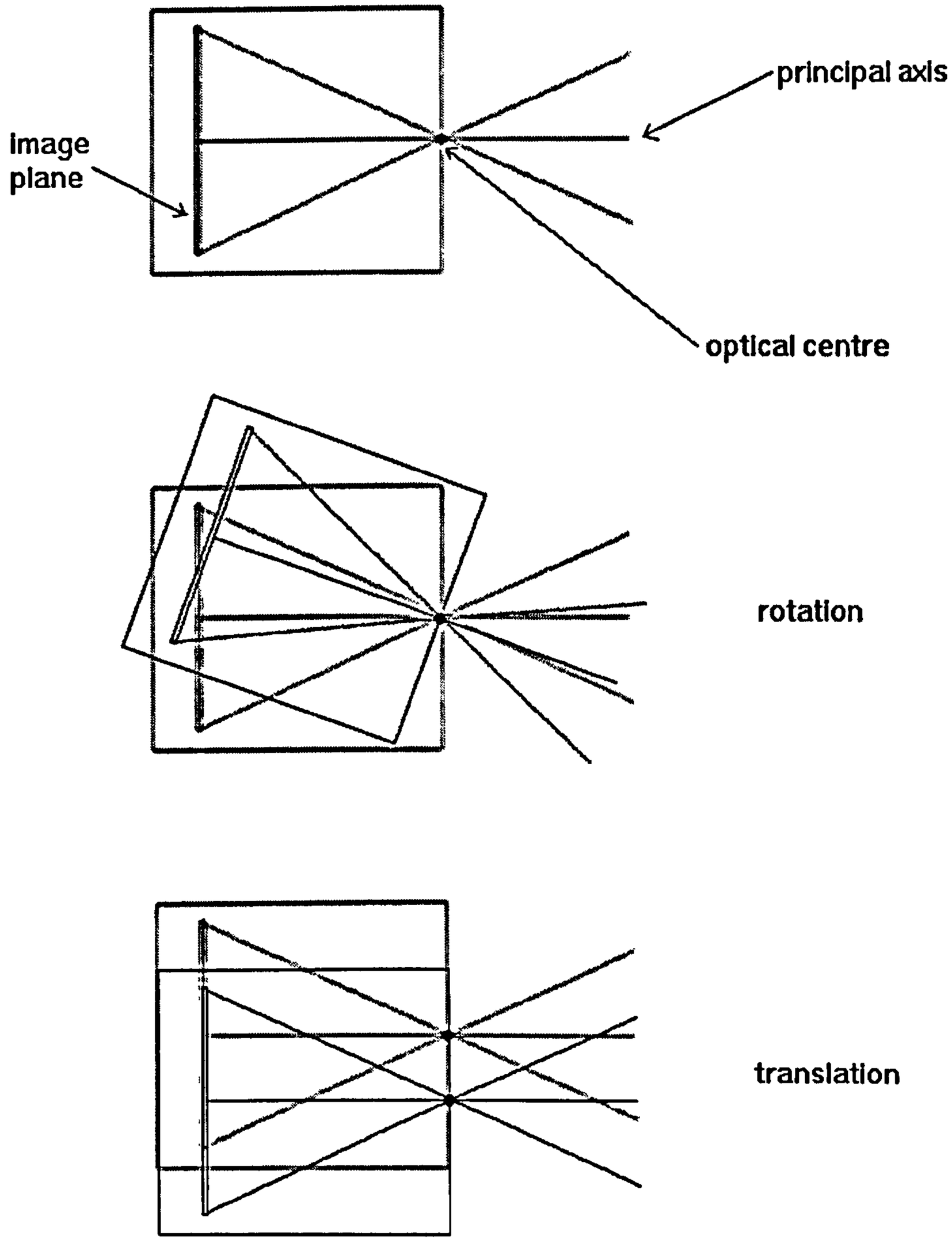

Figure 31 - rotation and translation

Camera alignment mountings are not easily or usually made in order to rotate a camera around its optical centre, so any attempted correction in angle usually introduces a translation. This makes the correct alignment of a camera a very difficult task using image-based methods. Pitch and yaw errors do produce keystone distortion, which could theoretically be detected, but the amount of this distortion is small enough for small rotations that it is difficult to produce a reliable calibration based upon its measurement. It is therefore important to separate the effects of rotation and translation on the camera system, so that it can be correctly aligned. In other words, rotations need to be measured independently of translations, and vice versa.

3.2.2 Calibration by using a distant target

One way to separate the errors is to use a distant target. When a target is photographed, the translation between the two cameras has a negligible effect on the images. Any difference in the images is due to a difference in rotation, assuming the focal lengths, principal points and other internal parameters are consistent, which can then be corrected for by rotating the cameras so that they produce identical images. Unfortunately, placing a calibration target at very large distances is not easily achieved within a laboratory setting. Imaging an outdoors landscape scene would be one way in which a target at effectively infinity could be realised, but such a scene is unlikely to be an optimal calibration target and will be difficult to match in an automated way. Even if an ideal calibration scene was available, the assumption that the internal parameters of each camera in a pair are the same rarely applies to real hardware.

It is possible to simulate the effect of a target at infinity by using parallel lines to determine vanishing points[51], however the accuracy is questionable (within 2 pixels per camera for Echigo's study[51], using a low-resolution camera), not only because of measured errors but also because of unmeasured errors arising from the small depth range of the points used for verification.

3.2.3 Separating treatment of lens and sensor

If the lens is removed, the camera system has far fewer internal parameters to consider. Principal distance and principal point are no longer relevant, and the only parameters left are the pixel size and pixel aspect ratio. Though the camera can no longer form an image, it is still possible to align the cameras based on the physical surface of their sensors. In this way, the sensors and lenses can be examined separately, eliminating the coupling between rotations and translations described above.

3.3 Laser diffraction from the sensor surface

3.3.1 A technique for finding the principal point of a digital camera

Clarke[47] shows a technique for determining the principal point of a digital camera by using a laser directed at its CCD to produce a reflected diffraction pattern. The type of diffraction pattern produced is shown below in Figure 32:

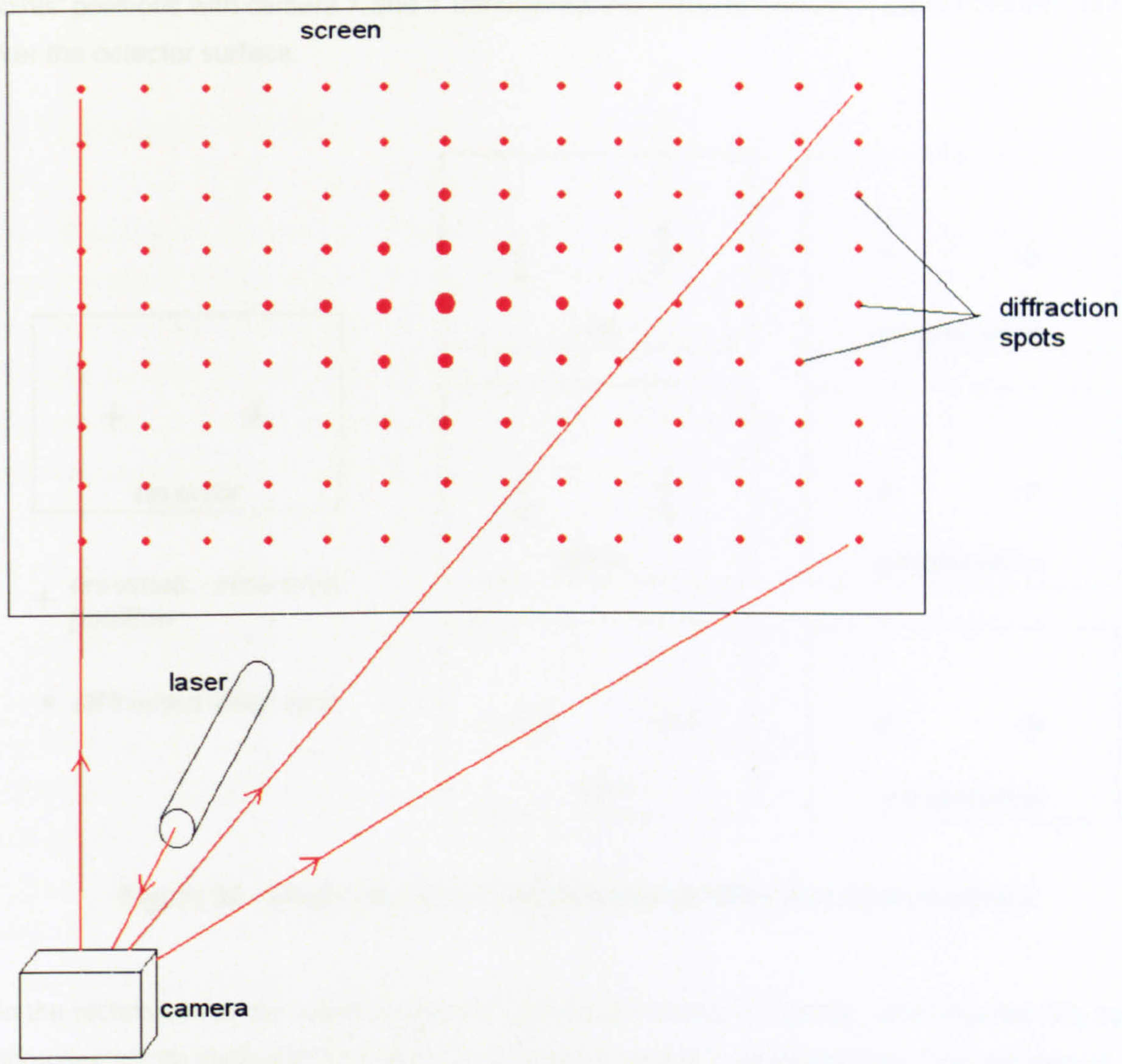

Figure 32 - diffraction pattern produced by a digital image sensor

The camera and beam are arranged so that the reflected (zero-order) beam is coincident with the incoming beam, and once the lens is attached, the principal point is the point where the beam strikes the detector.

3.3.2 Adapting Clarke's technique to align a digital camera

This technique can be adapted to align a digital camera absolutely in four dimensions, three of which are the three possible rotations. With the lens removed and the camera switched off, the diffraction pattern produced will rotate as the camera rotates. The pattern will roll as the camera rolls, it will rise as the camera pitches up, and move left and right as the camera yaws. It will also expand or shrink depending on the z-translation of the camera, but importantly the scaling effect can be entirely separated from the rotational effects. Also, the pattern is independent of horizontal and vertical positioning, and does not vary with translations in those directions, since CCDs are made with such accuracy (enough to produce no observable variation in diffracted

points' positions with camera X and Y translation) that the pixel divisions are of constant size over the detector surface.

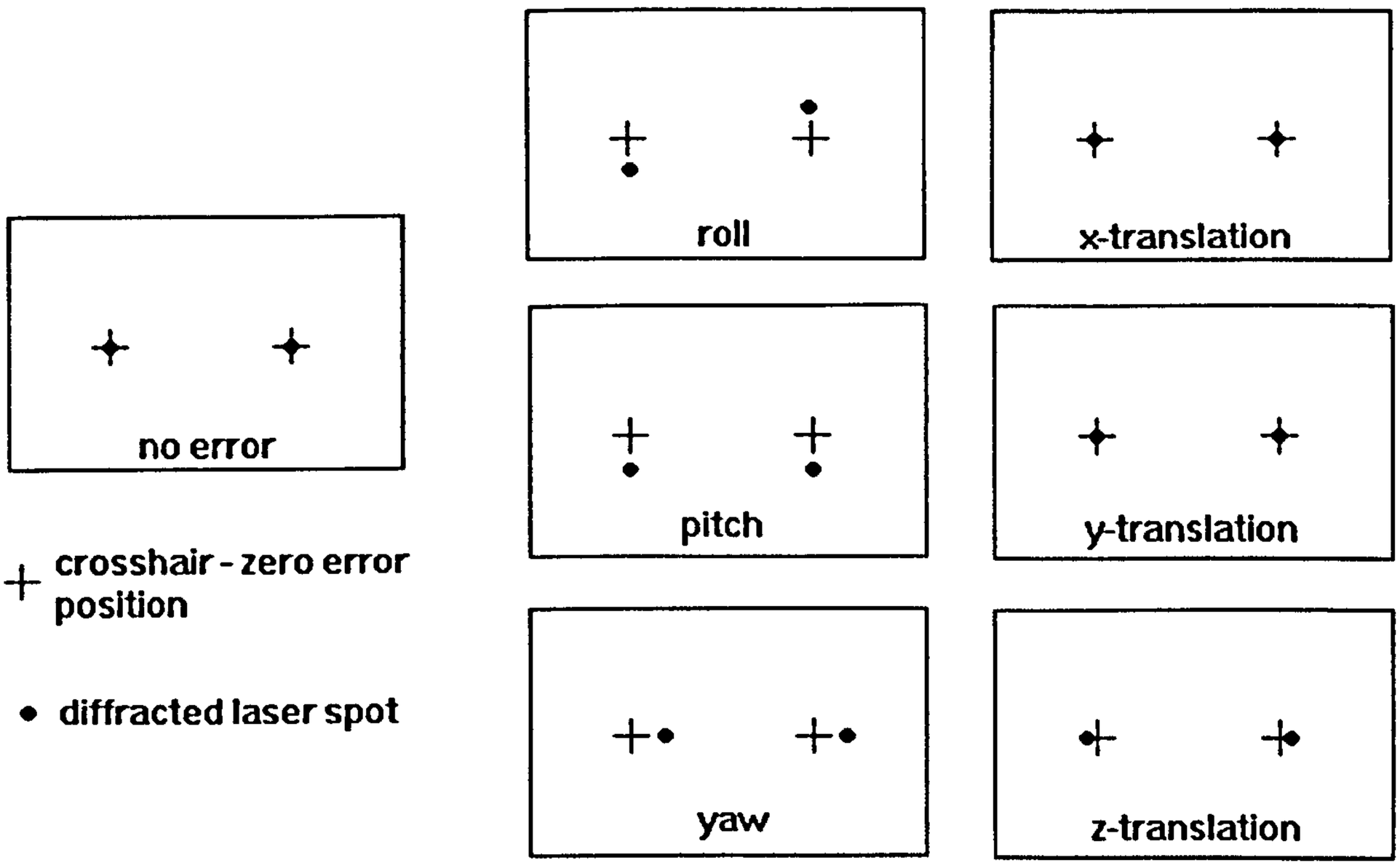

Figure 33 - effects of camera movements on diffracted spot positions

So the technique can be used to align the camera absolutely in rotation, and requires only two diffraction points each with an x and y coordinate to perform the calibration. Two degrees of freedom, the camera x and y positioning, are not addressed by this calibration, but may be estimated by locating the laser spot on the surface of the sensor. This may be performed by attenuating the laser and taking an image with the sensor.

3.4 Instability

3.4.1 Interior orientation stability problems

Digital cameras commonly suffer from poor interior orientation stability[26]. In some cameras the sensor may not be firmly fixed with respect to the camera body, the whole lens may move with respect to the camera body, and the lens elements may move differentially within the lens assembly[26]. These problems may make it difficult to perform precision calibrations with digital cameras, so choice of equipment plays a large part in determining how good a calibration can be. The Basler A113CPs chosen do not suffer from these problems[28].

3.4.2 Lens system instability

Lens systems are not typically completely rigid, because of tradeoffs involved in trying to make a system where the lenses are held in place relative to the camera sensor, but stresses in the materials are low[82]. Image shifts of several pixels are common when a lens system is knocked or shaken. Some of these shifts rectify themselves if the lens system is left alone, as accumulated stresses are released, but there can still be shifts of several pixels remaining. For this reason, lenses and their mountings should be handled with care.

3.4.3 Addressing problems with the lens mounting

Lenses may be able to be made more rigid than in their supplied state by means of a few simple adjustments. The Schneider Cinegon 1.4/8mm[30] lenses used were examined and made as robust as possible in the following ways. The front plate of the lens was removed, and the screws inside tightened as a measure to help reduce any internal movement in the lens assembly. The screws holding the back plate in place were also tightened, though this is probably the weakest link in terms of structural rigidity of the system.

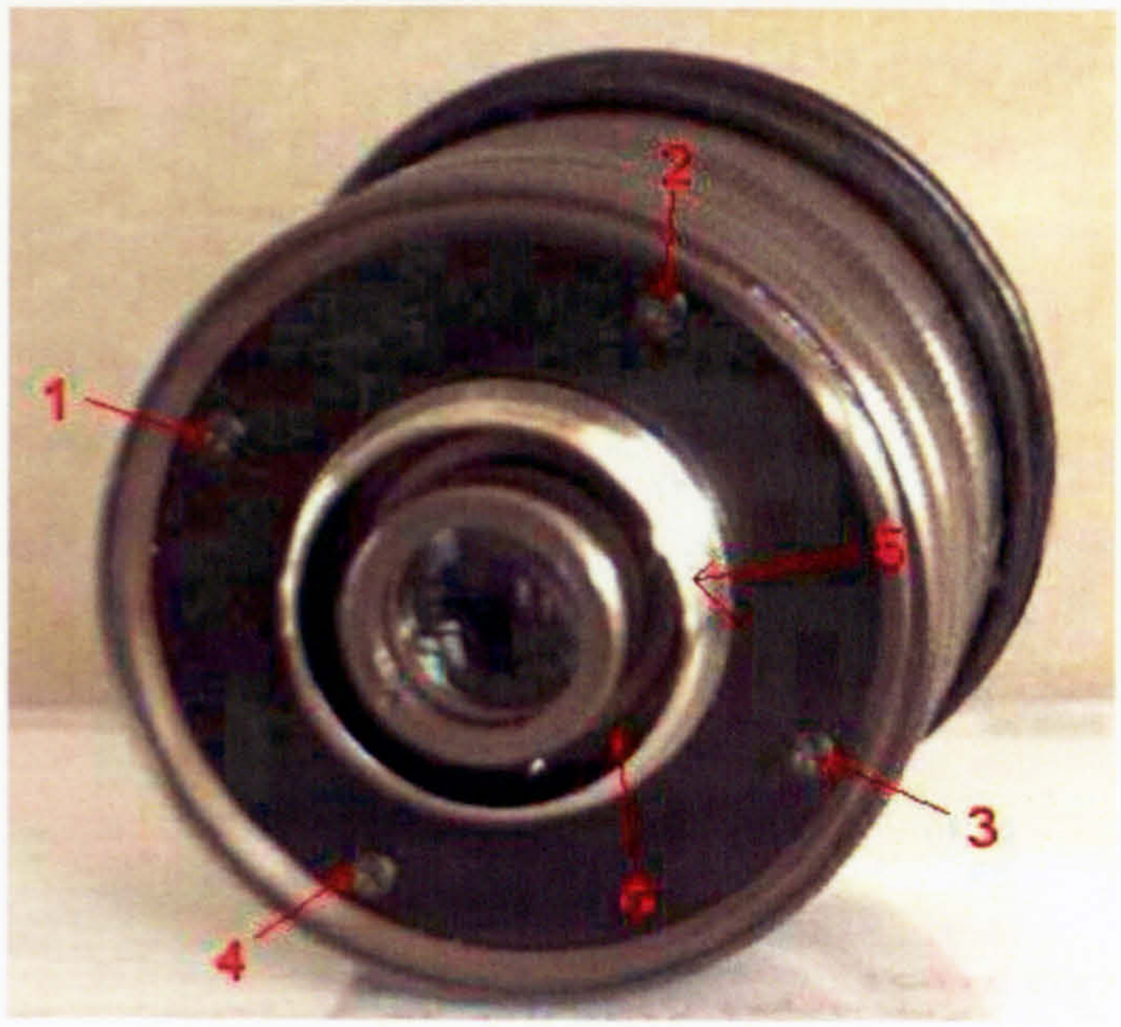

Figure 34 - the rear of the lens assembly, for a Schneider Cinegon 1.4/8mm lens

In Figure 34, points 1 to 4 are the screws holding the back plate to the interior lens mounting section, labelled 6. The C-mount adapter, marked as number 5, is only held in place by the

pressure from the back plate, and it is this area of the lens that is thought to be most susceptible to motion.

3.5 Summary

This chapter has shown how it is important to separate rotation and translation in calibrating a stereoscopic camera system, and describes a method by which this separation may be achieved. Lens instability, another of the serious challenges to an accurate camera calibration, is also discussed. By addressing these issues, an accurate new calibration method can be formulated.

In order to assess the usefulness of a camera calibration, the relationship between calibration parameters and image errors must be quantified. Chapter 4 models the behaviour of a two-head stereoscopic camera, and shows the calculation of this relationship.

4 Possible sources of error in the proposed scheme

4.1 Requirements

4.1.1 Method

Stereo theory usually considers a perfect pair of cameras. This refers to a situation where the only difference between the two cameras is their horizontal separation (in the case of parallel cameras). In the real world, there will be small differences between parameters of the cameras such as their orientation or focal lengths[83], and this will introduce errors into the stereo image pairs.

In order to find out how significant these errors are and how they will affect the camera design, the most critical image error must first be identified. This might be determined by the fusional requirements of the viewer as described in 1.6, which would mean that vertical disparity is perhaps the most sensitive error[69]. Alternatively, the desired errors might all be set as close to zero as permitted by the resolution of the camera (or display).

The errors in the stereo camera pair due to various alignment errors can be calculated. Camera orientation must be considered, as well as quantities such as principal distance and principal point. Each camera can be placed with six degrees of freedom, three translational and three rotational.

The approach taken is to model how an image point's location would change under a camera orientation change in each of the 6 degrees of freedom. A few simplifying assumptions are made. Lens distortion is not included in the analysis at this time, because this distortion cannot be easily corrected in hardware, for both horizontal and vertical disparities. Instead, low-distortion lenses[30] are chosen. The lenses have a distortion of only about 3% at the corner of the image for a 1284x1024 sensor with a pixel size of 6.7μm, for objects further away than 45cm. This distortion is quite low for a lens with such a wide (69° maximum diameter) field of view. The effects of distortion will be discussed later in 9.

The following section shows how these disparities were calculated, along with the calculation of disparities arising from principal distance and principal point errors. A slightly different approach is taken from Zhao et al[84] (which considers how some parameters affect depth estimates), whereby a more general treatment of errors and their effects is considered. Artificial disparity is calculated in both the vertical and horizontal directions for all nine errors. Equations from

Equation 7 to Equation 15 are referenced from published literature, while subsequent equations are independently derived.

4.1.2 Calculations of disparities produced by camera errors

A pinhole camera approach is used, as opposed to raytracing, since this result, while approximate, can be applied to any camera instead of just a single optical configuration. In this treatment, all angles are in radians. Pixel values are taken from the upper left of the detector with image x and y directions aligned with the world X and Y directions (with no error), while the origin of the world coordinate system is the optical centre (pinhole), which is also the centre of rotation and translation for the system. The directions are defined in Figure 35, which is a modification of Figure 25:

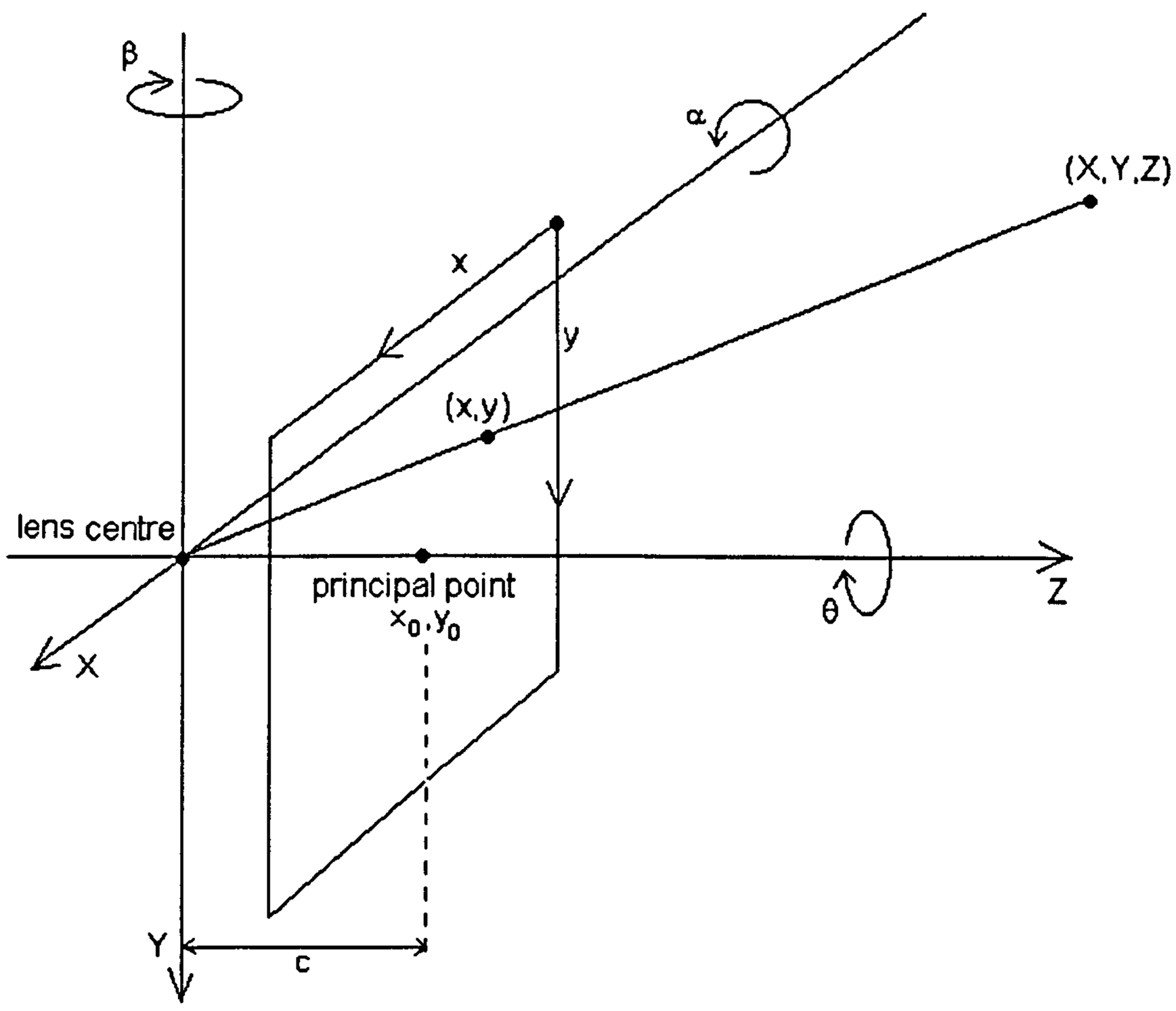

Figure 35 – definition of coordinate axes

The transformation between the real world and image space is given by[85]:

$$\gamma\begin{bmatrix}\underline{x}\\1\end{bmatrix} = \underline{P}\begin{bmatrix}\underline{X}\\1\end{bmatrix}$$

Equation 7

where:

$$\underline{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\underline{P} = \underline{C}.[\underline{R} \mid \underline{t}]$$

$$\underline{X} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

Equation 8

Here, x and y are image coordinates (measured in pixels from the upper left corner of the image array), X, Y and Z are world coordinates (i.e. the three-dimensional coordinates of space, centred at the optical centre/theoretical pinhole) in mm, $\underline{C}$ is the 3x3 calibration matrix, $\underline{R}$ is the 3x3 rotation matrix, and $\underline{t}$ is the 1x3 translation matrix. γ is a scaling factor (a scalar) and is calculated later in this chapter. $\underline{C}^{85}$ is defined below, with parameters subject to variation given an error Δ:

$$\underline{C} = \begin{bmatrix} (c + \Delta c).p_x & 0 & x_0 + \Delta x_0 \\ 0 & (c + \Delta c).p_y & y_0 + \Delta y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Equation 9

p_x and p_y are the number of pixels per mm in the x and y directions on the detector, x_0 and y_0 are the pixel coordinates of the principal point on the detector, c is the principal distance in mm, Δc is principal distance error in mm, Δx_0 is the principal point x-coordinate error in pixels, and Δy_0 is the principal point y-coordinate error in pixels.

4.1.3 Camera parameters

In this chapter, values for disparity and camera errors will be calculated, in addition to the presentation of general equations for their calculation. To produce these values, certain parameters need to be defined, and the values will apply to the particular camera to be built during the project. These values are:

$$
\begin{aligned}
c &= 8mm \\
p_x &= p_y = 149.3mm^{-1} \\
0 &\le x \le 1284 \\
0 &\le y \le 1024 \\
x_0 &= 642 \\
y_0 &= 512 \\
Z &= 1m
\end{aligned}
$$

Equation 10

x_0 and y_0 are estimates of the principal point position, using the centre of the sensor. Errors in these parameters are quantified later in sections 4.1.14 and 4.1.15.

4.1.4 Rotation matrices

The rotation matrices are given for roll, pitch and yaw in Equation 11, Equation 12 and Equation 13:[86]

$$\text{Roll, } \theta: \quad \underline{R_\theta} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Equation 11

$$\text{Pitch, } \alpha: \quad \underline{R_\alpha} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$$

Equation 12

$$\text{Yaw, } \beta: \underline{R_\beta} = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$$

Equation 13

These matrices can be combined into a single rotation matrix in an order-independent way, valid for small angles, giving[87]:

$$\underline{R} = \begin{bmatrix} 1 & -\theta & \beta \\ \theta & 1 & -\alpha \\ -\beta & \alpha & 1 \end{bmatrix}$$

Equation 14

The translation matrix is:

$$\underline{T} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix}$$

Equation 15

Where ΔX, ΔY and ΔZ are expressed in mm.

4.1.5 Results

Multiplying out the matrices produces:

No error ($\theta = \alpha = \beta = \Delta X = \Delta Y = \Delta Z = \Delta c = \Delta x_0 = \Delta y_0 = 0$):

$$x = x_0 + c.p_x.\frac{X}{Z}$$
$$y = y_0 + c.p_y.\frac{Y}{Z}$$

Equation 16

which is the same as Equation 4 shown earlier. With errors, denoting the image coordinates with error x' and y':

$$x' = x_0 + \Delta x_0 + \frac{(c + \Delta c).p_x}{\gamma}[X - Y\theta + Z\beta + \Delta X]$$

$$y' = y_0 + \Delta y_0 + \frac{(c + \Delta c).p_y}{\gamma}[X\theta + Y - Z\alpha + \Delta Y]$$

Equation 17

where γ is calculated from Equation 7 as:

$$\gamma = -X.\beta + Y.\alpha + Z + \Delta Z$$

Equation 18

So:

$$\Delta x = x' - x$$

$$\Delta y = y' - y$$

$$\Delta x = \Delta x_0 + c.p_x.\left(\frac{\left(1 + \frac{\Delta c}{c}\right).(X - Y.\theta + Z.\beta + \Delta X)}{-X.\beta + Y.\alpha + Z + \Delta Z} - \frac{X}{Z}\right)$$

$$\Delta y = \Delta y_0 + c.p_y.\left(\frac{\left(1 + \frac{\Delta c}{c}\right).(X.\theta + Y - Z.\alpha + \Delta Y)}{-X.\beta + Y.\alpha + Z + \Delta Z} - \frac{Y}{Z}\right)$$

Equation 19

4.1.6 Per-pixel vertical disparity measures

In the following sections, expressions for the disparities due to the nine specified errors are given, assuming in each case that errors other than the one under consideration are zero. Values for the maximum errors in each case, for an artificial vertical disparity of one pixel, are calculated. Some measures are dependent on principal distance, and all are calculated using Equation 19. The pixel pitch of 149.3pixels/mm quoted in Equation 10 is based upon a pixel size of 6.7μm, the size of the pixels in the cameras to be used here. Where disparity equations are reduced, Equation 16 is used to substitute for world coordinates, so that disparities are expressed (as far as possible) as functions of image coordinates.

4.1.7 Roll

Using Equation 19 a find disparities due to roll gives:

$$\Delta x = -\frac{c.p_x.Y.\theta}{Z}$$
$$\Delta y = \frac{c.p_y.X.\theta}{Z}$$

Equation 20

Using Equation 16, this simplifies to:

$$\Delta x = -(y - y_0).\theta$$
$$\Delta y = (x - x_0).\theta$$

Equation 21

The vertical disparity varies with y, and is largest at the left and right edges of the image, as shown in Figure 36. For values as given in Equation 10, for vertical disparity to be less than one pixel (in either direction) across the image requires that $|\theta| < 0.0895°$.

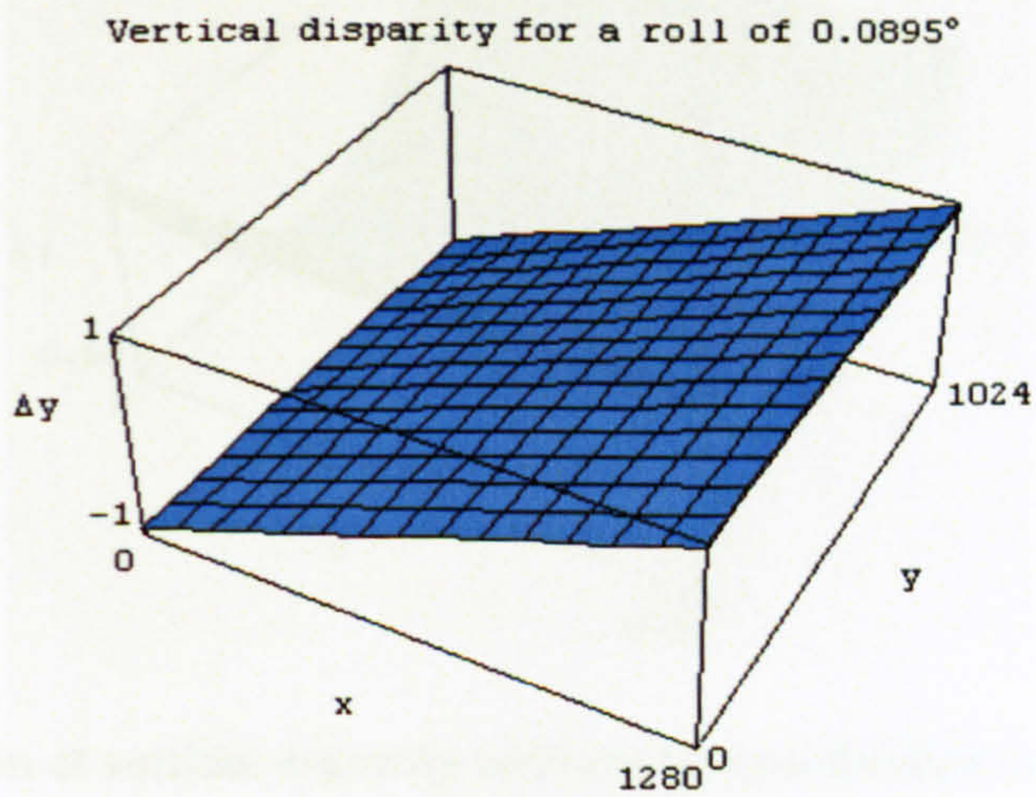

Figure 36 - variation of vertical disparity with image coordinates for a 0.0895° roll

4.1.8 Pitch

Using Equation 19 to find disparities due to pitch gives:

$$\Delta x = -\frac{c.p_x.X.Y.\alpha}{Z.(Z + Y.\alpha)}$$

$$\Delta y = -\frac{c.p_y.(Y^2 + Z^2).\alpha}{Z.(Z + Y.\alpha)}$$

Equation 22

Using Equation 16, this simplifies to:

$$\Delta x = -\frac{\alpha.(x - x_0).(y - y_0)}{c.p_y + \alpha.(y - y_0)}$$

$$\Delta y = -\frac{\alpha.(c^2.{p_y}^2 + (y - y_0)^2)}{c.p_y + \alpha.(y - y_0)}$$

Equation 23

The vertical disparity varies with y and principal distance, and is largest at the top and bottom of the image, as shown in Figure 37. For values as given in Equation 10, for vertical disparity to be less than one pixel (in either direction) across the image requires that $|\alpha| < 0.0405°$.

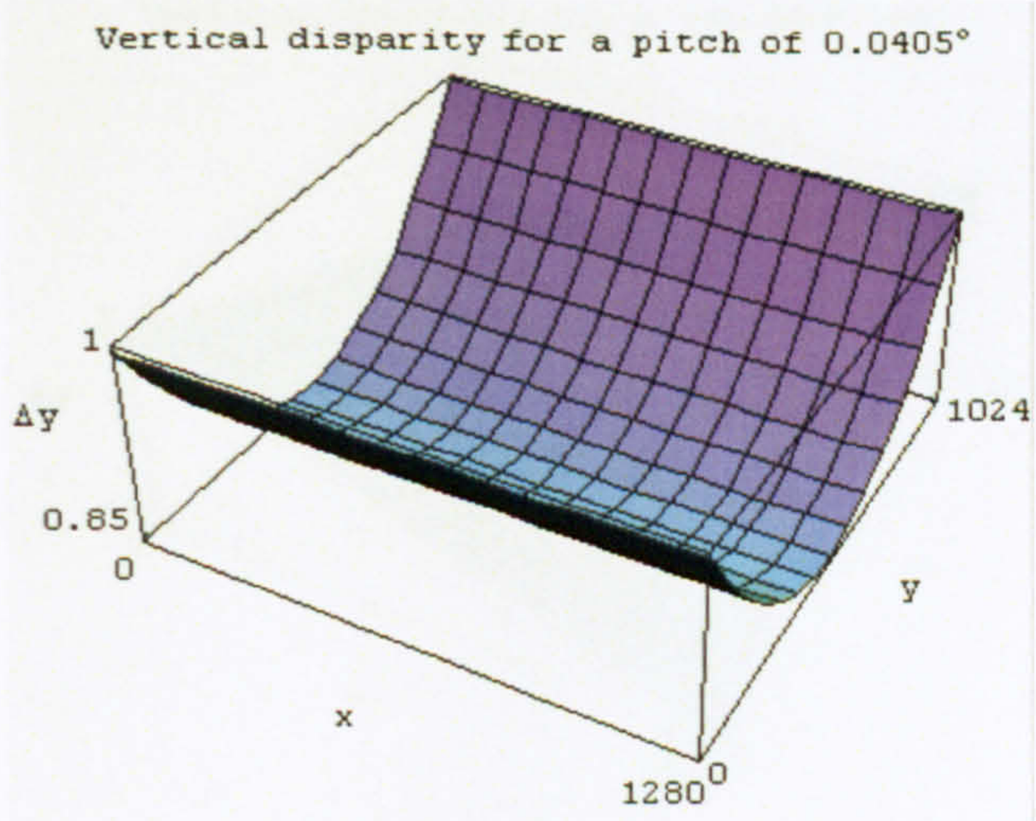

Figure 37 - variation of vertical disparity with image coordinates for a 0.0405° pitch at 8mm principal distance

4.1.9 Yaw

Using Equation 19 to find disparities due to yaw gives:

$$\Delta x = \frac{c.p_x.\left(X^2 \quad Z^2\right).\beta}{Z.(Z - X.\beta)}$$

$$\Delta y = \frac{c.p_y.X.Y.\beta}{Z.(Z - X.\beta)}$$

Equation 24

Using Equation 16, this simplifies to:

$$\Delta x = \frac{\beta.\left(c^2.{p_x}^2 \quad (x - x_0)^2\right)}{c.p_x \quad \beta.(x - x_0)}$$

$$\Delta x = \frac{\beta.(x - x_0).(y - y_0)}{c.p_x \quad \beta.(x - x_0)}$$

Equation 25

The vertical disparity varies with x, y and principal distance, and is largest at the corners of the image, as shown in Figure 38. For values as given in Equation 10, for vertical disparity to be less than one pixel (in either direction) across the image requires that $|\beta| < 0.209°$.

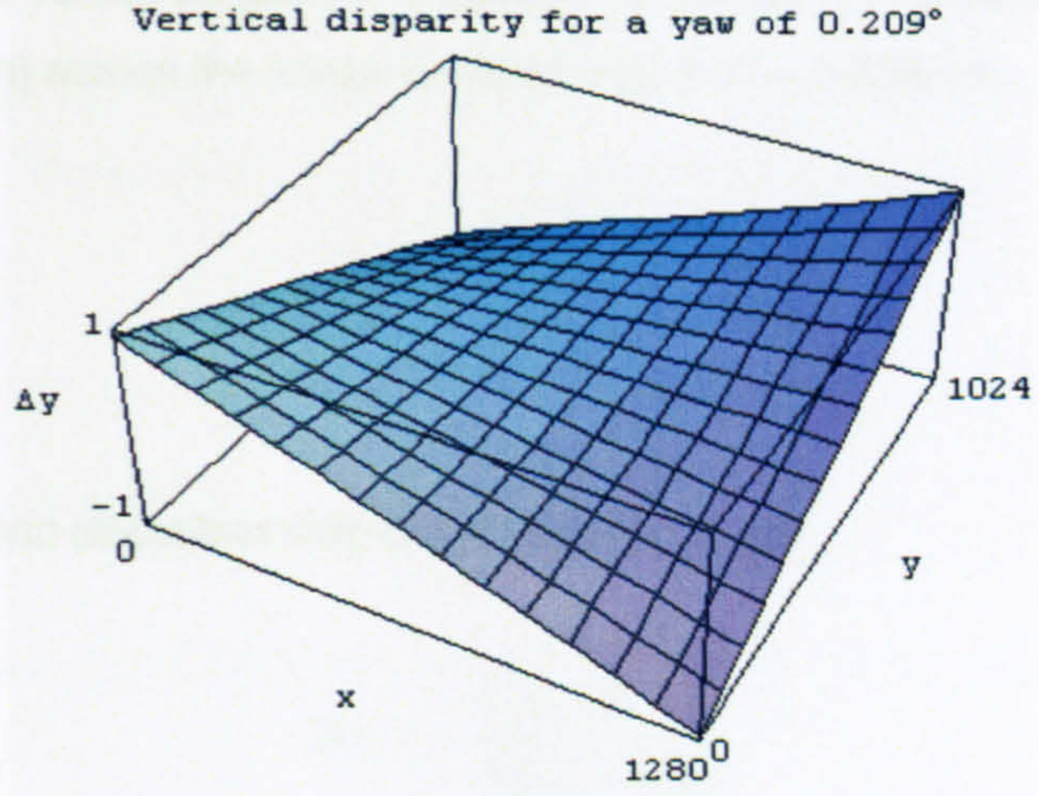

Figure 38 - variation of vertical disparity with image coordinates for a 0.209° yaw at 8mm principal distance

4.1.10 X-translation

Using Equation 19 to find disparities due to X-translation gives:

$$\Delta x = \frac{c.p_x.\Delta X}{Z}$$

$$\Delta y = 0$$

Equation 26

The vertical disparity is zero for this error.

4.1.11 Y-translation

Using Equation 19 to find disparities due to Y-translation gives:

$$\Delta x = 0$$

$$\Delta y = \frac{c.p_y.\Delta Y}{Z}$$

Equation 27

The vertical disparity varies with principal distance and Z-coordinate, and is the same over the whole image array. For values as given in Equation 10, for vertical disparity to be less than one pixel (in either direction) across the image requires that $|\Delta Y| < 0.838mm$.

4.1.12 Z-translation

Using Equation 19 to find disparities due to Z-translation gives:

$$\Delta x = -\frac{c.p_x.X.\Delta Z}{Z.(Z+\Delta Z)}$$

$$\Delta y = -\frac{c.p_y.Y.\Delta Z}{Z.(Z+\Delta Z)}$$

Equation 28

Using Equation 16, this simplifies to:

$$\Delta x = \frac{(x - x_0)\Delta Z}{Z + \Delta Z}$$
$$\Delta y = \frac{(y - y_0)\Delta Z}{Z + \Delta Z}$$

Equation 29

The vertical disparity varies with y and the Z-coordinate, and is largest at the top and bottom of the image, as shown in Figure 39. For values as given in Equation 10, for vertical disparity to be less than one pixel (in either direction) across the image requires that $|\Delta Z| < 1.96$mm.

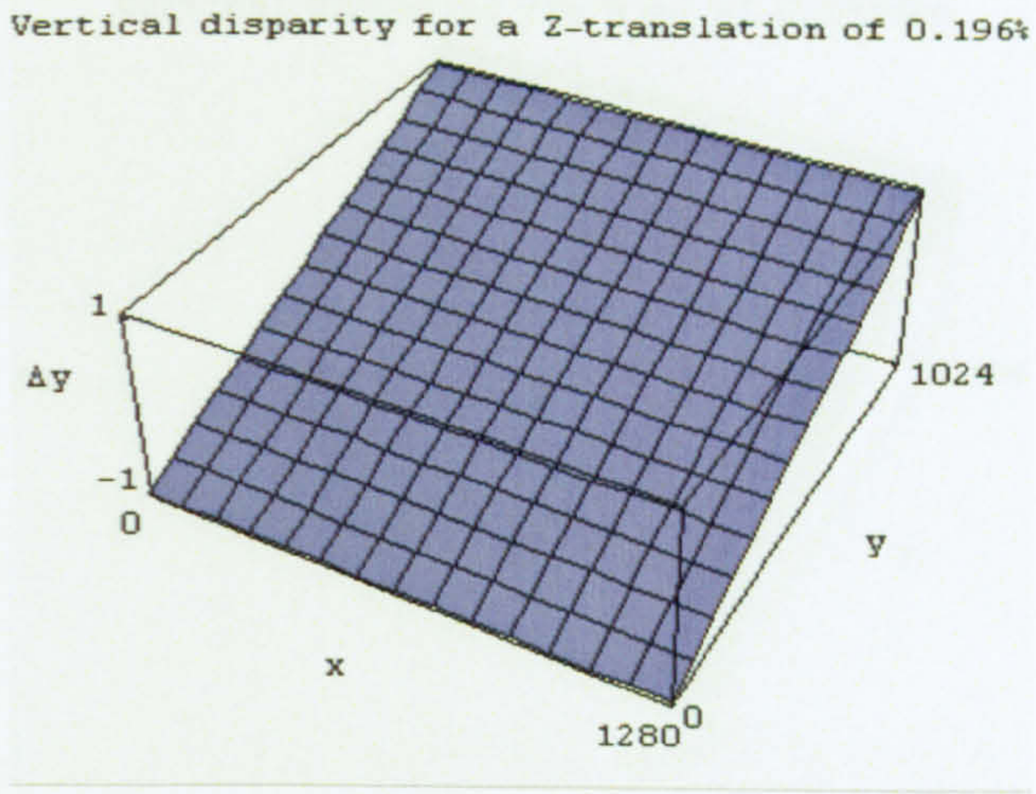

Figure 39 - variation of vertical disparity with image coordinates for a 0.196% Z-translation, equal to 1.96mm at 1m Z-distance

4.1.13 Principal distance change

Using Equation 19 to find disparities due to principal distance change gives:

$$\Delta x = \frac{p_x . X . \Delta c}{Z}$$
$$\Delta y = \frac{p_y . Y . \Delta c}{Z}$$

Equation 30

Using Equation 16, this simplifies to:

$$\Delta x = (x - x_0)\frac{\Delta c}{c}$$

$$\Delta y = (y - y_0)\frac{\Delta c}{c}$$

Equation 31

The vertical disparity varies with y and is largest at the top and bottom of the image, as shown in Figure 40. For values as given in Equation 10, for vertical disparity to be less than one pixel (in either direction) across the image requires that $|\Delta c| < 0.0156$mm.

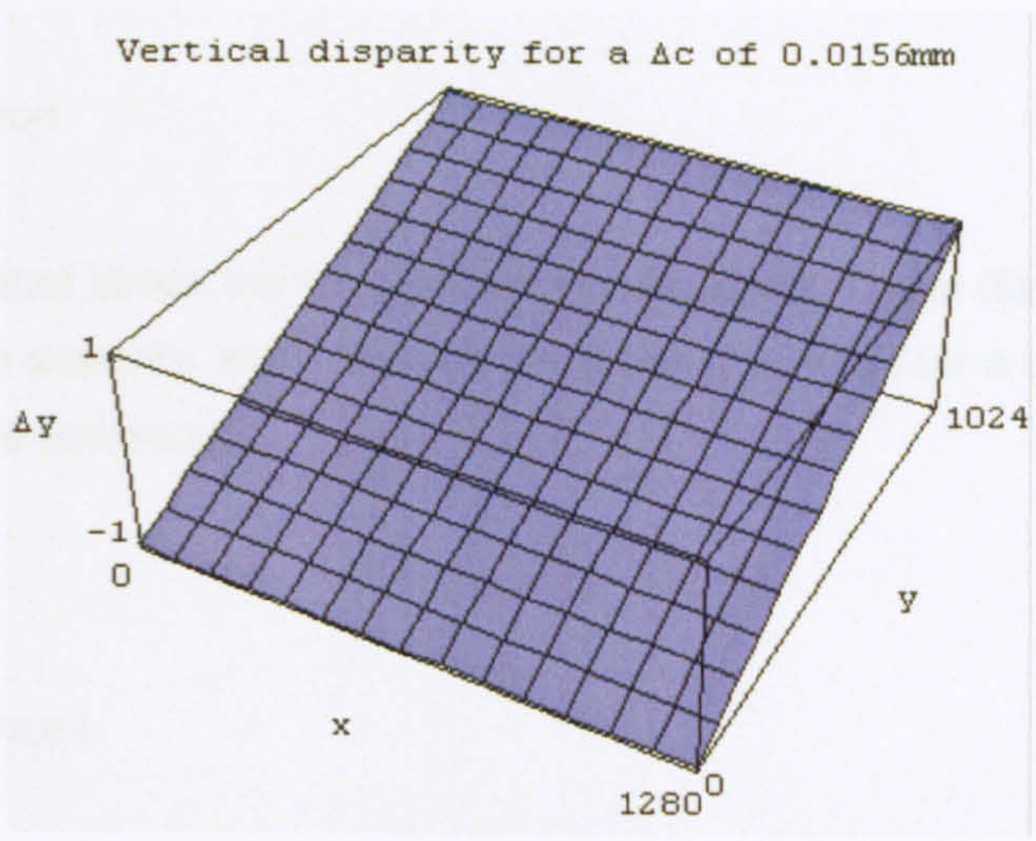

Figure 40 - variation of vertical disparity with image coordinates for a 0.0156mm principal distance error at an 8mm principal distance

4.1.14 X-centre translation

From Equation 19 and Equation 16:

$$\Delta x = \Delta x_0$$
$$\Delta y = 0$$

Equation 32

The vertical disparity caused by x-centre translation is zero.

4.1.15 Y-centre translation

From Equation 19 and Equation 16:

$$\Delta x = 0$$
$$\Delta y = \Delta y_0$$

Equation 33

The vertical disparity caused by y-centre translation is equal to the y-centre translation. For vertical disparity to be less than one pixel (in either direction) across the image requires that $|\Delta y_0| < 1$ pixel.

4.1.16 Error combination

The disparities calculated above are for each error individually. These disparities combine to form the overall image disparity, and once the errors are quantified for a calibrated camera, the overall disparity can be estimated.

4.1.17 Slidebar tolerances

The above sections describe the errors on a single camera. There is a second type of error, which can be called a "slidebar error" because of its relevance to a single camera stereo system using a slidebar to translate between images. This is where there is a systematic error which is the same in both cameras, relative to their separation. There are only two slidebar errors which can produce an artificial disparity, which are slidebar roll and slidebar yaw. There are no slidebar analogues of the other seven errors applicable to single cameras.

Slidebar roll is shown below in Figure 41:

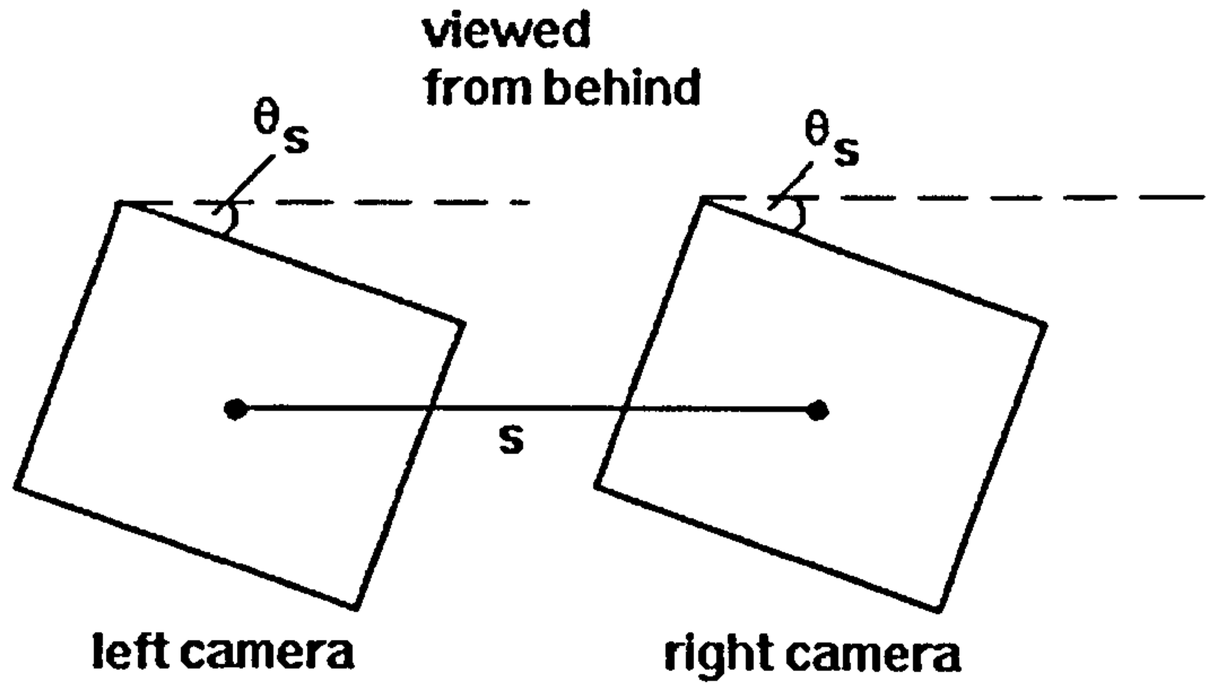

Figure 41 - slidebar roll

The two cameras are separated by a distance s, and the slidebar roll is of magnitude θ_s. The slidebar roll produces an effective single-camera translation in the X and Y directions, the effect of which is as described in sections 4.1.10 and 4.1.11. These are:

$$\Delta X = s.(1 - \cos\theta_s)$$
$$\Delta Y = s.\sin\theta_s$$

Equation 34

Substituting using Equation 26 and Equation 27 gives:

$$\Delta x = \frac{c.p_x.s.(1-\cos\theta_s)}{Z}$$

$$\Delta y = \frac{c.p_y.s.\sin\theta_s}{Z}$$

Equation 35

The vertical disparity varies with the separation, principal distance and the Z-coordinate. For values as given in Equation 10, for vertical disparity to be less than one pixel (in either direction) across the image requires that $|\theta_s| < 0.480°$.

Slidebar yaw is shown below in Figure 42:

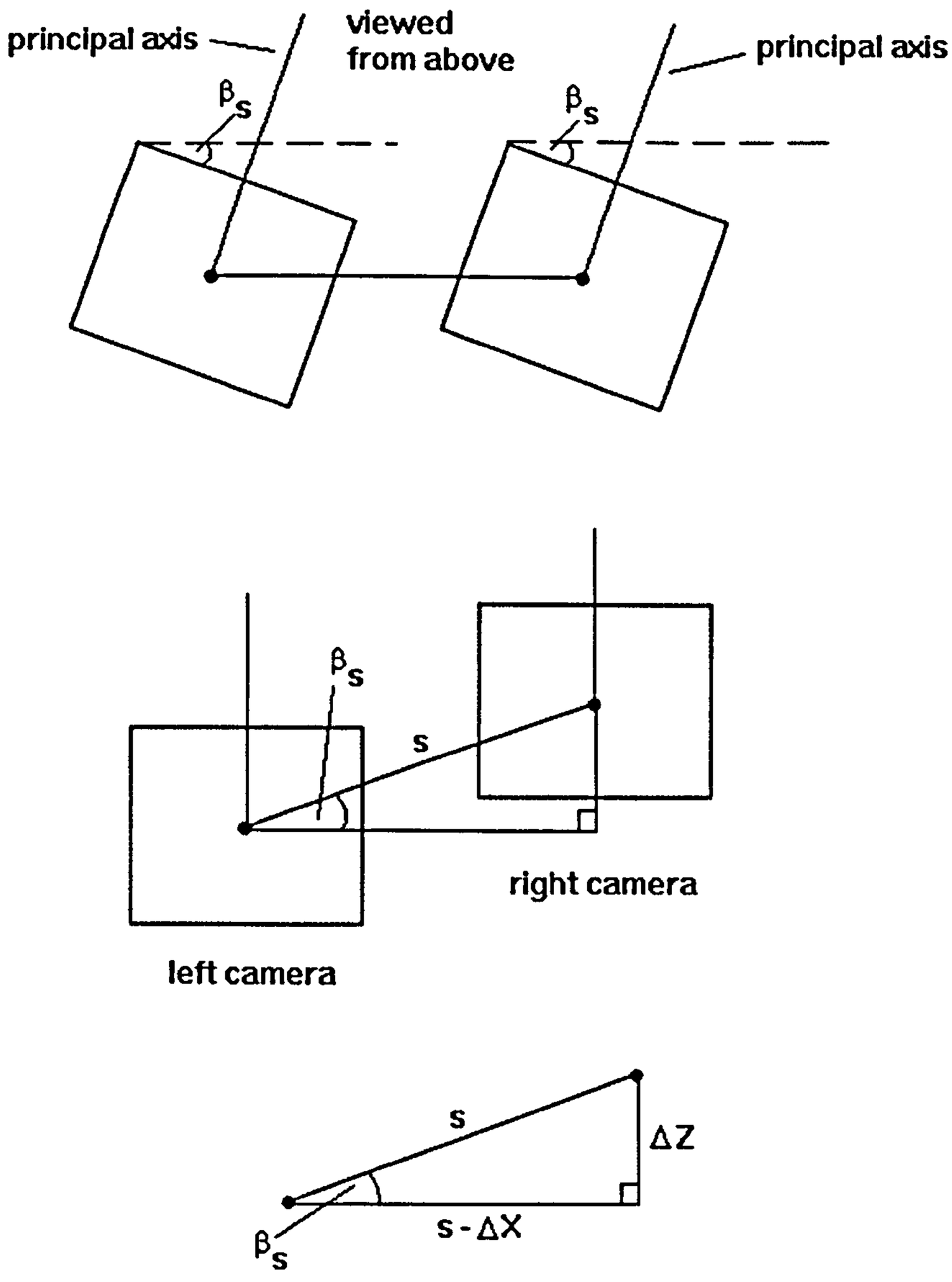

Figure 42 - slidebar yaw

The two cameras are separated by a distance s, and the slidebar yaw is of magnitude β_s. The slidebar yaw produces an effective single-camera translation in the X and Z directions, the effect of which is as described in 4.1.10 and 4.1.11. These are:

$$\Delta X = s.(1 - \cos \beta_s)$$
$$\Delta Z = s. \sin \beta_s$$

Equation 36

Substituting using Equation 26 and Equation 29 gives:

$$\Delta x = \frac{c.p_x.s.(1 - \cos \beta_s)}{Z} + \frac{s.(x - x_0).\sin \beta_s}{Z + s.\sin \beta_s}$$
$$\Delta y = \frac{s.(y - y_0).\sin \beta_s}{Z + s.\sin \beta_s}$$

Equation 37

The vertical disparity varies with separation, y and the Z-coordinate, and is largest at the top and bottom of the image. For values as given in Equation 10, for vertical disparity to be less than one pixel (in either direction) across the image requires that $|\beta_s| < 1.12°$.

These measures show that errors which are identical for both cameras have a lesser effect on the stereoscopic image than errors which occur only on one camera.

4.1.18 Temporal synchronisation

The cameras must be synchronised so that moving objects in the field of view must not appear significantly displaced relative to the static scene between the two images. The cameras can be synchronised using the trigger unit to better than 0.1ms[88], so for values as defined in Equation 10, an object moving at 480°/s relative to the camera will move one pixel over that 0.1ms. To put this speed into a linear context, an object moving perpendicular to the principal axis has to move at 8.375m/s per metre of distance from the camera in order to move more than one pixel in the 0.1ms. For almost all normal situations, the scene being imaged won't contain objects moving as fast at such distances as that, so the camera synchronisation can be assumed to be sufficient for the application.

4.2 Stability requirement

4.2.1 Strength and rigidity

The camera pair needs to be held together in some kind of mounting apparatus which fixes their position and orientation with respect to each other. The mounting needs to be stable enough to maintain the alignment so that movement of the mounted cameras does not introduce image errors which are significant. The measures for the stability of the mounting are exactly as described above, and will contribute additional disparity terms if the mounting is not stable. The fundamental requirements are that the mounting should hold each camera tightly so that the cameras cannot move significantly within the mounting itself, and also that the mounting should not bend significantly - such movements could introduce rotational and translational errors.

4.2.2 Thermal stability

Thermal stability is also important, because digital cameras generate heat while being operated, and this will cause expansion of the material around them. The greatest camera movement due to thermal expansion will be in the direction with the greatest length of material, which is the horizontal separation. Translations in this direction cause no vertical disparity, but do cause horizontal disparity.

Thermal expansion is calculated using Equation 38[89]:

$$\Delta X = \alpha.L.\Delta T$$

Equation 38

where α is the coefficient of linear thermal expansion, L is the length over which the expansion occurs, and ΔT is the change in temperature. Combining Equation 26 with Equation 38 gives Equation 39 for calculating the effect of thermal expansion:

$$\Delta x = \frac{c.p.\alpha.L.\Delta T}{Z}$$

Equation 39

Assuming the camera mounting is made from aluminium, the value of α[89] is $2.4 \times 10^{-5} K^{-1}$. With other parameters as defined as given in Equation 10, a one-pixel horizontal disparity requires a

temperature change of 349K. This is much greater than any temperature rise likely to be encountered, and thermal expansion is therefore disregarded as a significant source of image error.

4.3 Summary

This chapter has calculated the relationship between camera calibration parameters and errors in images produced by a stereoscopic camera. This provides an indication of the level of accuracy in the calibration parameters that is required to produce images fit for human stereoscopic viewing, and allows image errors to be determined from calibration parameter errors.

In the following chapters, a method for calibrating a stereoscopic camera according to the principles in this thesis is described. The first part of this calibration procedure is the calibration of the sensor orientation, and is described in chapter 5.

5 Alignment of the camera sensor

5.1 Introduction

This chapter describes the method of alignment of a camera body containing a digital image sensor.

5.2 gives the theory behind the laser alignment technique, and describes the formation of the diffraction pattern from the sensor surface. As the sensor is moved in either roll, pitch, yaw or the Z-axis, the diffraction pattern's size, orientation and position changes, and these changes are calculated. Combining this with the calculations from 4 gives the relationship between image error in the camera and diffraction pattern change. The accuracy with which the diffraction pattern can be manipulated is tested, and evaluated in the context of the limits on image error this produces.

5.3 details the procedure of the laser alignment technique, as applied to a stereoscopic camera pair. The process is based upon defining a known coordinate system in space relative to the various components of the camera system, and aligning the cameras' sensors to the principal axis direction this defines. Errors introduced by this procedure are considered and quantified, to produce an estimate of the total image error caused by the laser alignment. This error is evaluated and compared with the image error due to aligning the camera pair based on the geometry of the camera bodies, to demonstrate the benefits of this procedure.

5.2 The laser alignment technique

5.2.1 The laser diffraction pattern

The size of a diffraction pattern from a regular grid can be calculated using the diffraction grating equation[89]:

$$n\lambda = d.\sin\theta$$

Equation 40

Where n is the diffraction order (positive integers), λ is the wavelength of the laser used, d is the spacing of the grid, and θ is the angular spacing between adjacent diffraction maxima.

For a helium-neon laser in air, the wavelength is 632.8nm. The colour filter array[40] of the sensor in the Basler A101CP camera[28] has near zero transmission at this wavelength in the G and B filters, and approximately 80% transmission in the R filters. This arrangement is shown below in Figure 43.

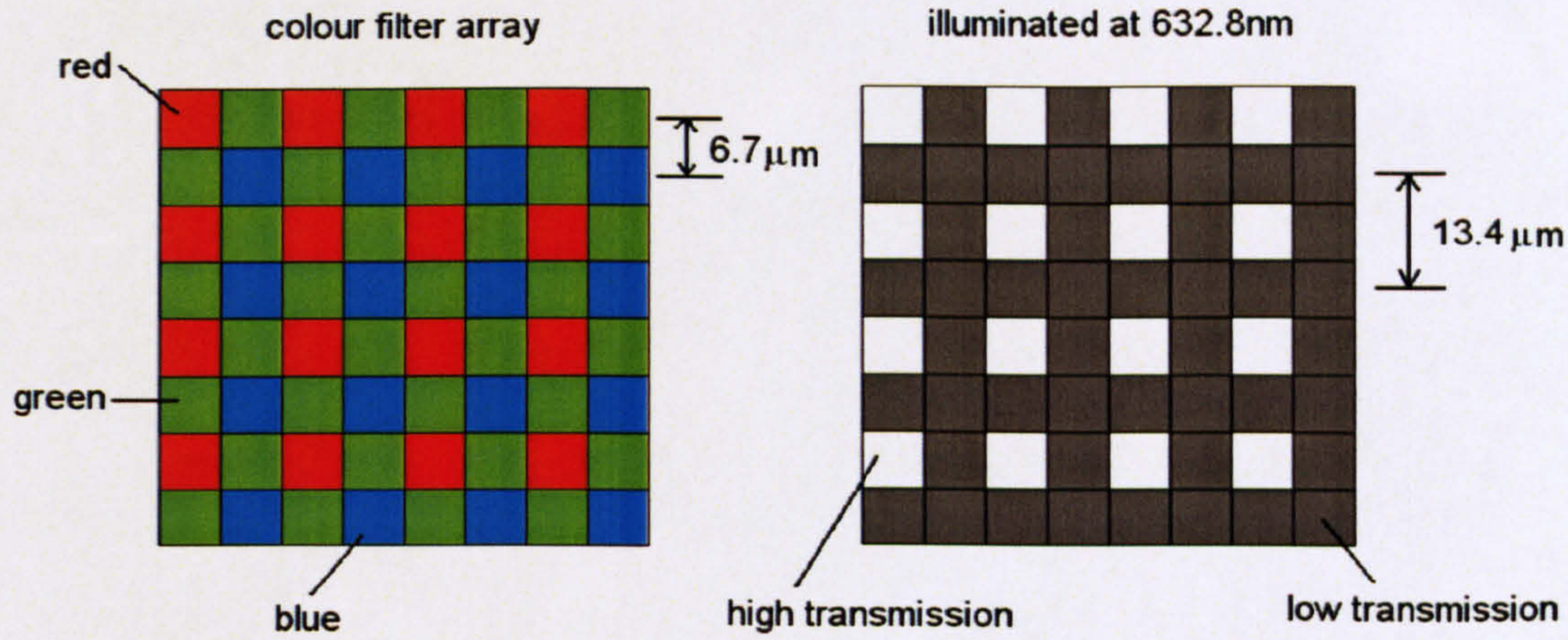

Figure 43 - diffraction of a HeNe laser beam from the colour filter array

The transmission varies in a grid pattern with a spacing d equal to two pixels, or 13.4μm. This forms the diffraction grid.

So the chip will produce a pattern with an angular spacing between diffraction maxima of about 2.7°. This spacing will conveniently give a grid spacing of a few centimetres on a screen at around a metre away (4.73cm at 1m). The pattern produced is shown in Figure 44, and Figure 45 shows the effect of the apparatus partially obstructing the diffraction pattern. Each spot appears approximately radially symmetric, and has a width of approximately 4mm.

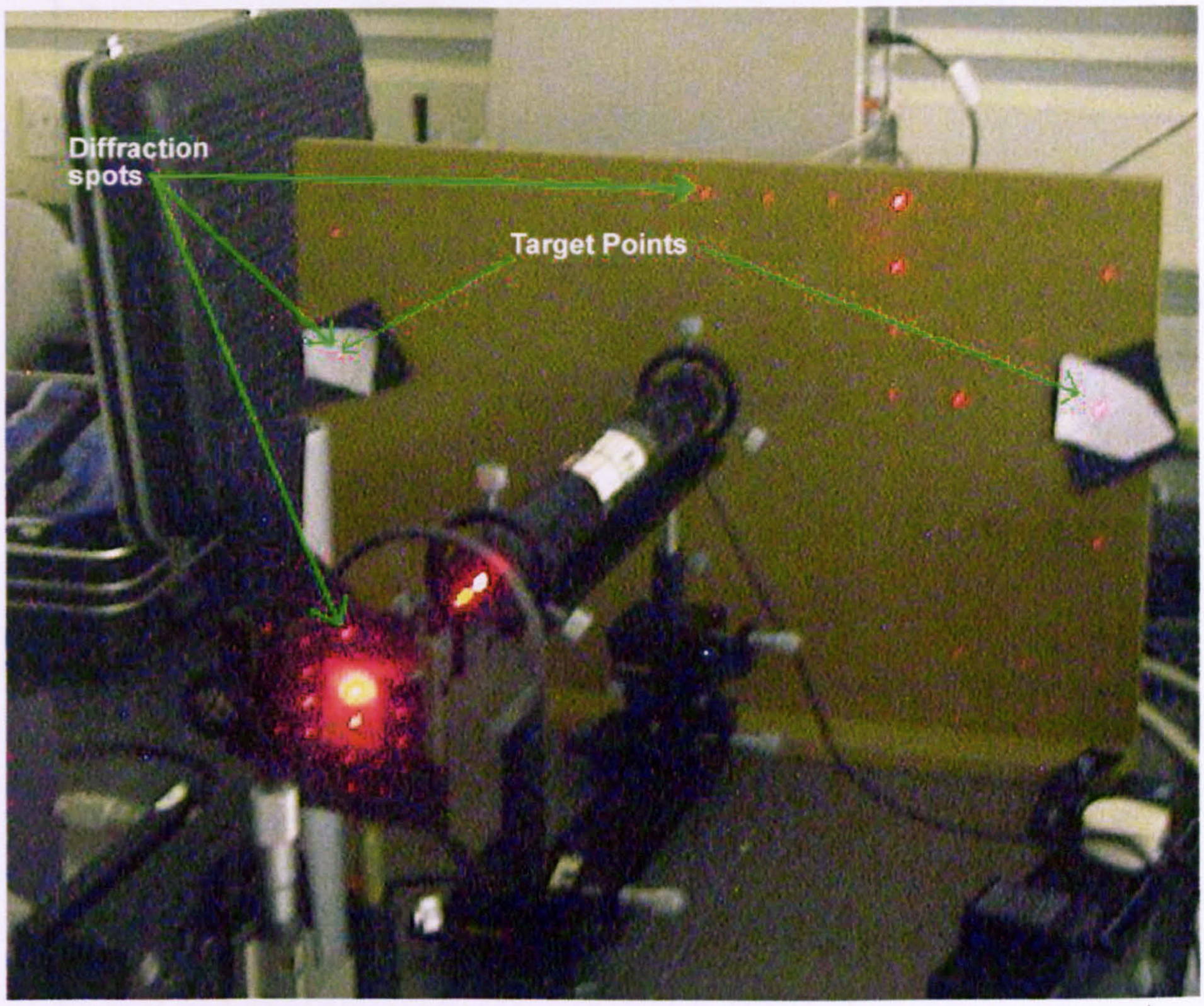

Figure 44 – the diffraction pattern from the CCD, with green arrows marking the diffracted spots. Many of the spots are blocked by parts of the apparatus before reaching the screen, but the grid pattern is clearly visible. The two target points as described in 3.3.2 are also marked.

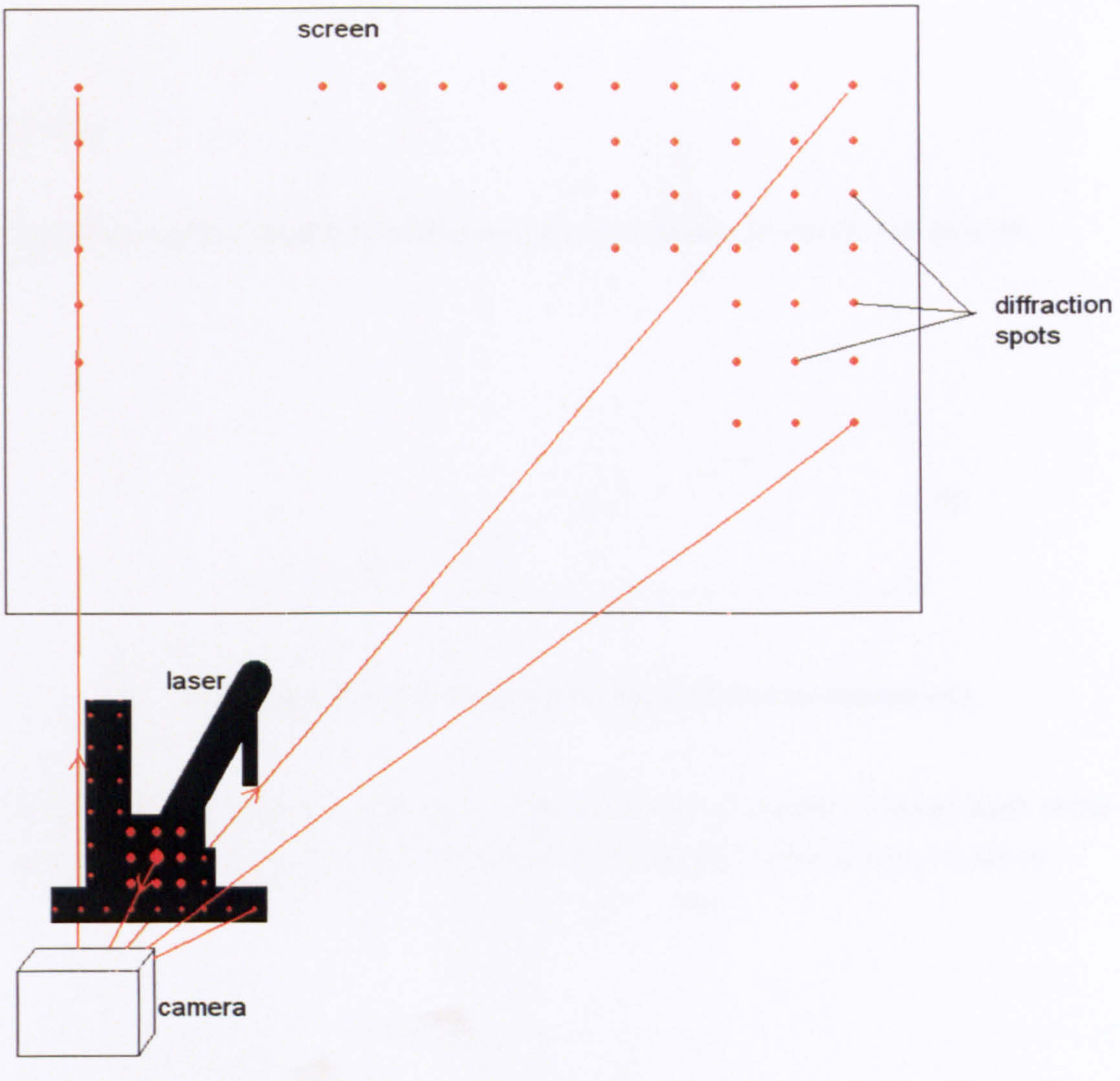

Figure 45 - schematic diagram to show the apparatus casting a shadow on the distant screen

This screen is set up at approximately 90cm from the sensor plane, and the spot spacing at the screen is about 4.5cm. The screen does not need to be precisely aligned to any of the equipment, since all that will be required from the apparatus is that the spot positions are the same for CCDs with identical orientations.

5.2.2 Required accuracy

The laser spots will move as the camera moves in all three angular directions, as well as during a translation in the z-direction. The motion of the spots can be calculated for a given movement of the camera. Below are spot movements calculated for a horizontally diffracted spot at 27cm from the centre of the pattern, with the screen 90cm from the camera. Motion amounts are calculated for a one pixel disparity (not simply a vertical disparity, but a disparity in any direction).

5.2.3 Roll

The pattern is simply rolled by the same angle as the camera, as shown in Figure 46.

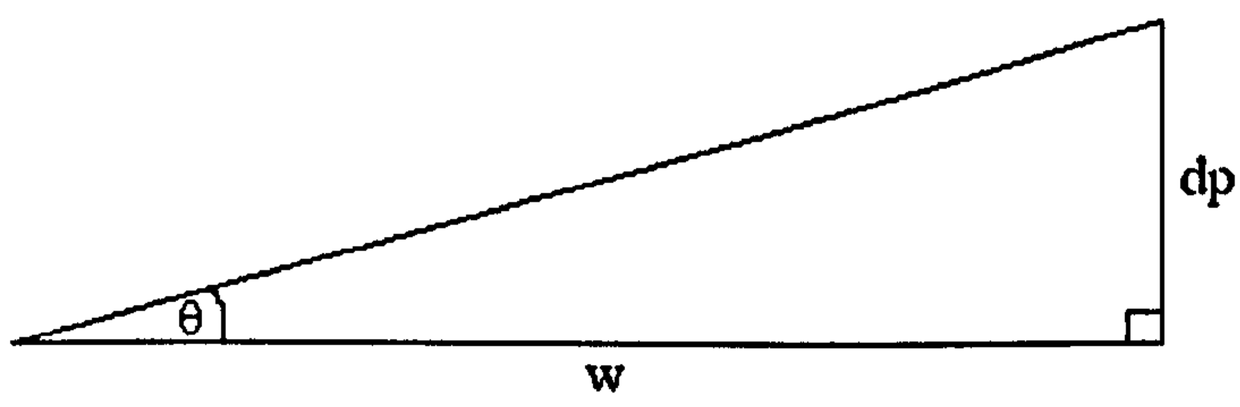

Figure 46 - diffraction spot movement dp due to camera roll

So for a one pixel movement, section 4.1.7 shows us that $\theta = 0.0895°$. The half-width of the pattern w = 27cm (using the sixth-order maxima), so the spot movement dp = 0.42mm.

5.2.4 Pitch and yaw

The pattern is pitched or yawed by twice the angle as the camera, since the pattern is reflected, as shown in Figure 47.

Figure 47 - diffraction spot movement dp due to camera pitch

For a one pixel movement, section 4.1.8 shows us that $\alpha = 0.0405°$ for an 8mm focal length. The distance to the screen l = 80cm, so the spot movement dp = 1.27mm. The movement is the same in magnitude for camera yaw, only in a perpendicular direction.

5.2.5 Z-translation

The size of the pattern changes as the camera is translated along the Z-axis, because the beams are diffracted at a constant angle from the sensor surface. The pattern grows in proportion to the change in distance, so that the change in pattern half-width, dp, for a pattern of half-width w, with the screen a distance l from the sensor, is, for a change in screen distance of ΔZ:

$$dp = \frac{w.\Delta Z}{l}$$

Equation 41

For a pattern of half-width w = 27cm, and a screen distance l = 90cm, a one pixel movement, which from section 4.1.12 is 1.96mm, requires that the spot movement dp is 0.59mm.

5.2.6 Accuracy of the alignment procedure

The positioning of the camera is controlled by an alignment stage, which uses micrometer stages to adjust the camera in the X, Y, Z, roll pitch and yaw directions. The accuracy of the alignment method was tested by attempting to repeat an initial spot position after z-translating and rotating in all three directions the camera, and measuring the average errors on the micrometers. The yaw and pitch micrometers were 50mm away from the axes around which they rotated, and the roll micrometer was 40mm away. With a screen at approximately 80cm from the camera, and a pattern half-width of approximately 30cm (from centre to side), the average micrometer errors over 15 trials were measured and pixel uncertainties calculated as shown:

Direction	Micrometer uncertainty/mm	Parameter uncertainty	Pixel uncertainty/pixels (main axis)	Spot location error/mm
Roll	0.003	0.0043°	0.048 (y)	0.02
Pitch	0.005	0.0057°	0.141 (y)	0.16
Yaw	0.003	0.0034°	0.085 (x)	0.10
Z-axis	0.17	0.17mm	0.087 (x)	0.06

Table 1 - estimated uncertainties per camera in the laser alignment procedure

Pixel errors caused by the errors in alignment are shown to be significantly less than one pixel in both axes, and the method is therefore suitable for aligning the camera system. Spot location

error varies with the type of measurement, and is highest for the two cases (pitch and yaw) where spot positions must be repeated from memory. For the two cases (roll and Z-axis) where spot positions are made equal across both visible target points, the errors are smaller, because memory of more than a couple of seconds need not be relied upon. Upper positioning error limits of 0.2mm are made in the first case, and 0.1mm in the second.

5.3 Alignment of the cameras' sensors

5.3.1 The laser and its mounting

The camera pair mounts on a pair of rail sliders, glued to an optical rail (a Newport PRL-36[90]). The rail edge serves as a definition of the X-direction in the apparatus, and the camera's detectors are aligned relative to this. A beam with zero yaw is defined as being perpendicular to the X-direction. The rail's top surface is parallel[90] to the optical bench on which the apparatus is aligned, and this plane defines a plane in which a beam has zero pitch. It also defines the plane in which two horizontally-diffracted beams fall which have zero roll angle. If co-reflection of the zero order beam is to be used to show zero camera pitch and yaw, the laser must be aligned so that the incoming beam has zero pitch and yaw.

The laser (a Hughes 3225H-PC, a helium-neon laser with a power output of 10mW) is mounted upon an optical rail (hereafter designated the laser rail), with the beam axis parallel to the long axis of the laser rail (a Melles griot model 07 ORN 007)[91]. The mounting attaches at two points on the laser body, each one having a fine position adjustment and a translation stage in the base. This allows the laser to be aligned parallel to the laser rail. This alignment is achieved by sliding a target back and forth along the laser rail, and adjusting the mount so that the laser spot stays at a constant position (to within 0.1mm) on that target. The setup for this procedure is shown below in Figure 48. The laser must be powered up for at least 30 minutes before alignment, since the beam alignment changes as the laser warms up. In front of the laser, an attenuating filter (which was a variable mirror, but could have been a cross-polariser or neutral density filter) is mounted so that the intensity of the beam can be controlled.

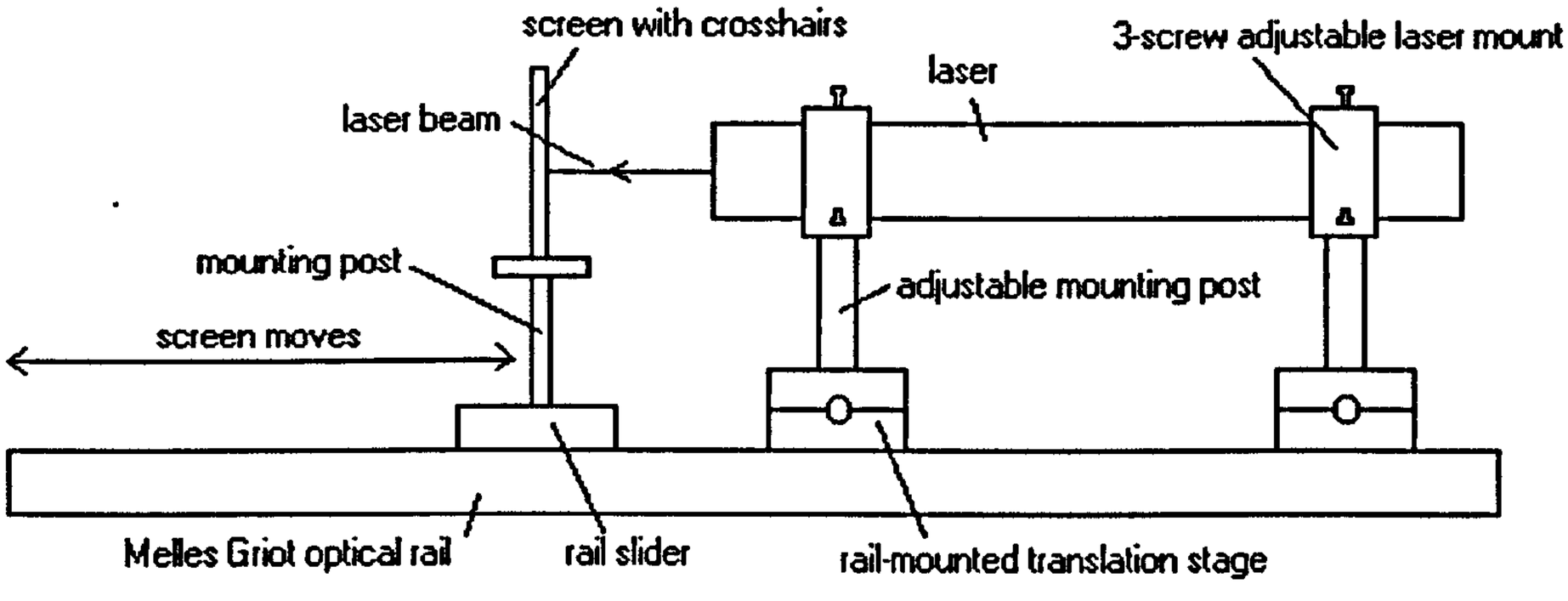

Figure 48 - laser alignment setup, side view

5.3.2 The camera and its alignment apparatus

The camera rail (shown in Figure 49) is a Newport optical rail[90], and sits on top of a raised bench mount, to position the camera in the path of the laser beam. The camera rail is orientated perpendicular to the laser rail by butting it up against a precision made 90° L-piece (90° ±0.05°) before locking it down to the raised mount through clearance holes. This procedure ensures that the camera rail is perpendicular to the laser beam, to the accuracy of the L-piece.

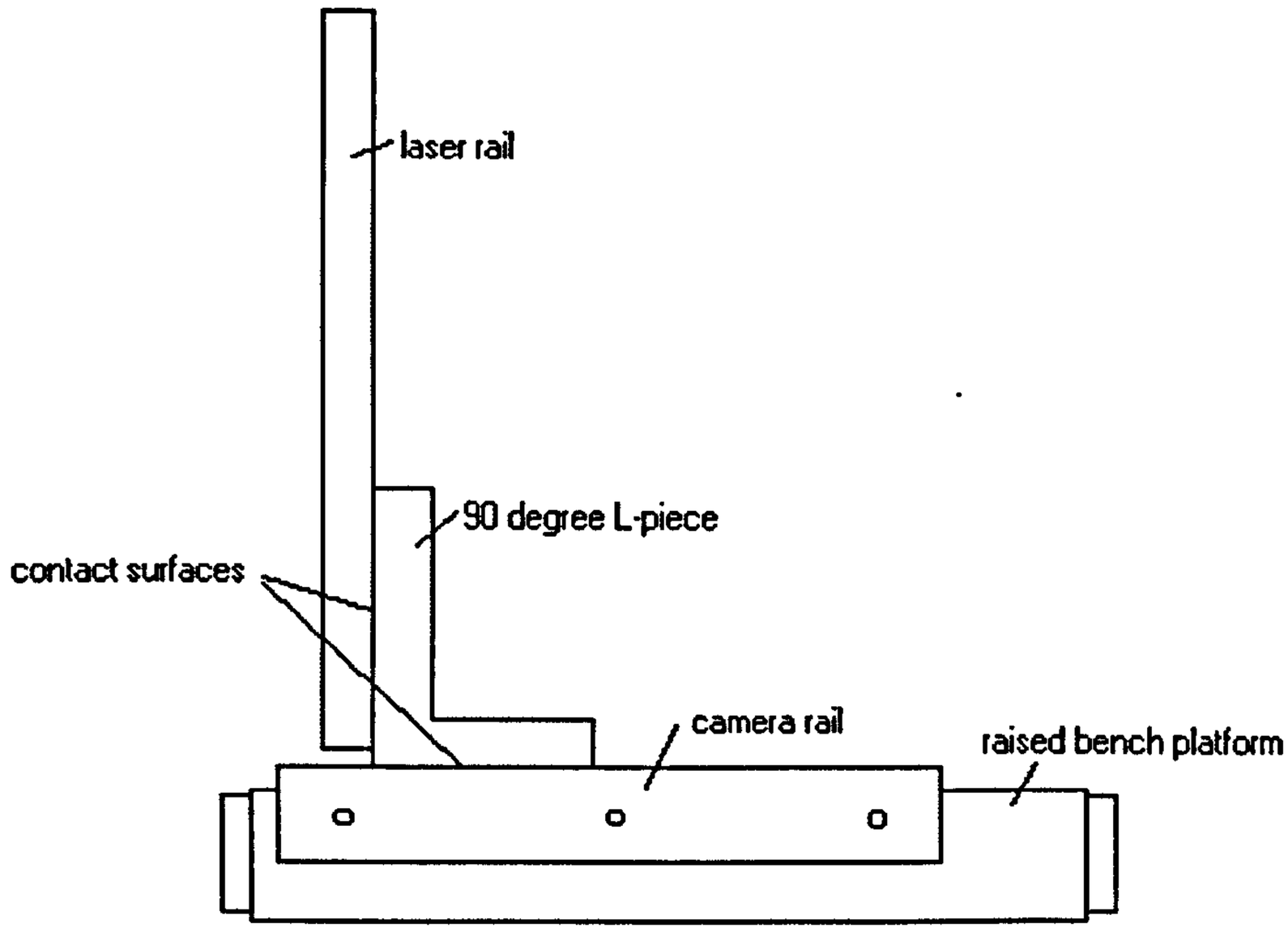

Figure 49 - aligning the camera rail, from above

The six-axis mount for aligning the camera sits on top of the camera rail, mounted on a rail slider[92].

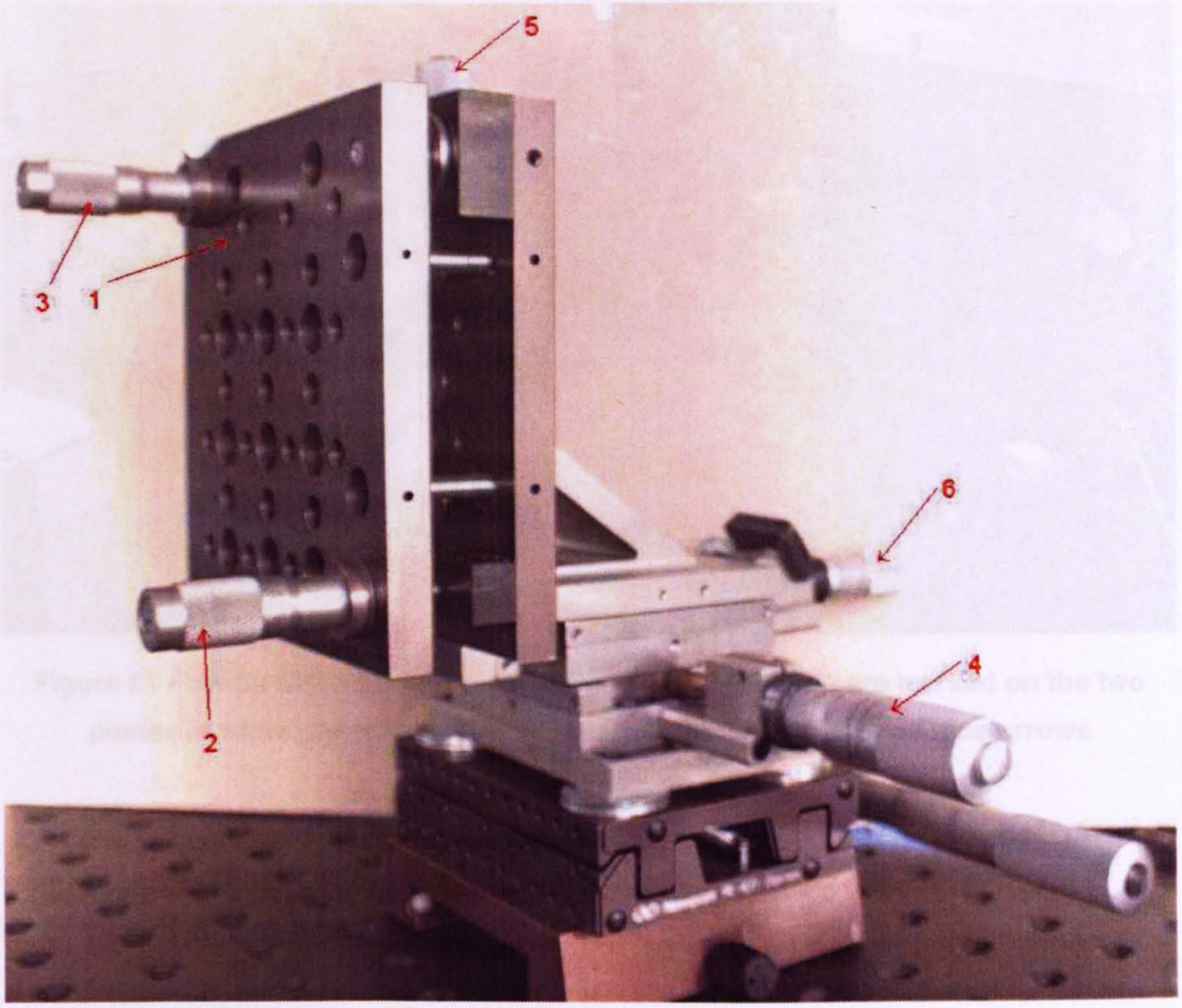

Figure 50 - the six-axis mount

The mount has micrometer screws for adjustment in the roll, pitch and yaw directions (labelled 1, 2 and 3 in Figure 50, the roll micrometer is obscured) as well as the X, Y and Z directions (labelled 4,5, and 6 in Figure 50). Attached to the alignment stage is a T-piece which has a pair of holes to which the camera is attached, facing forwards with the lens removed.

5.3.3 The target plane

The target plane is a fibreboard screen fixed to the bench behind the laser. The laser beam reflects and diffracts from the surface of the CCD, and produces a diffraction pattern on the screen, 80cm away. Two spots from this pattern are set to fall on two crosshairs in the target plane, show in Figure 51 below.

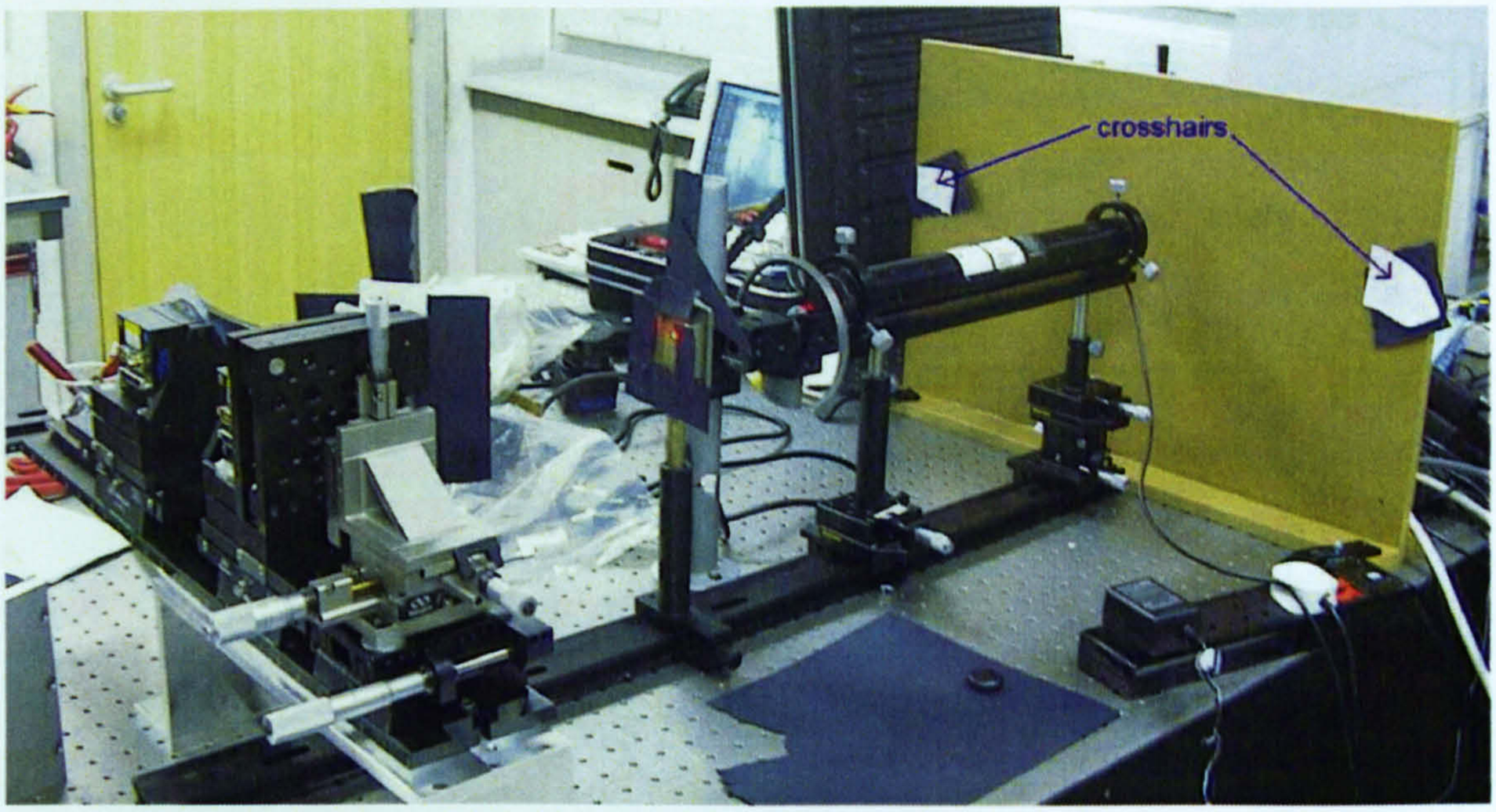

Figure 51 - bench setup for laser calibration. The crosshairs are marked on the two pieces of white paper affixed to the screen, marked by the two blue arrows

5.3.4 First camera alignment

The first camera is aligned so that the CCD is perpendicular to the laser beam, and therefore parallel to the laser rail. First, the camera is translated to make the laser beam hit the approximate centre of the CCD. Next, a pinhole screen on the laser rail is aligned with the incoming beam, and the surrounding diffraction pattern is used to align the camera in roll, pitch and yaw. The camera is then more accurately aligned in pitch and roll by sliding a target at a fixed height around the bench, and ensuring that the height of the horizontally diffracted beams are in a plane parallel to the bench surface, and at the same height as the incoming beam. The crosshairs in the target plane are adjusted so that a pair of horizontally diffracted points falls on them, and these are set as reference points for future calibrations. Figure 52, below, shows the alignment apparatus.

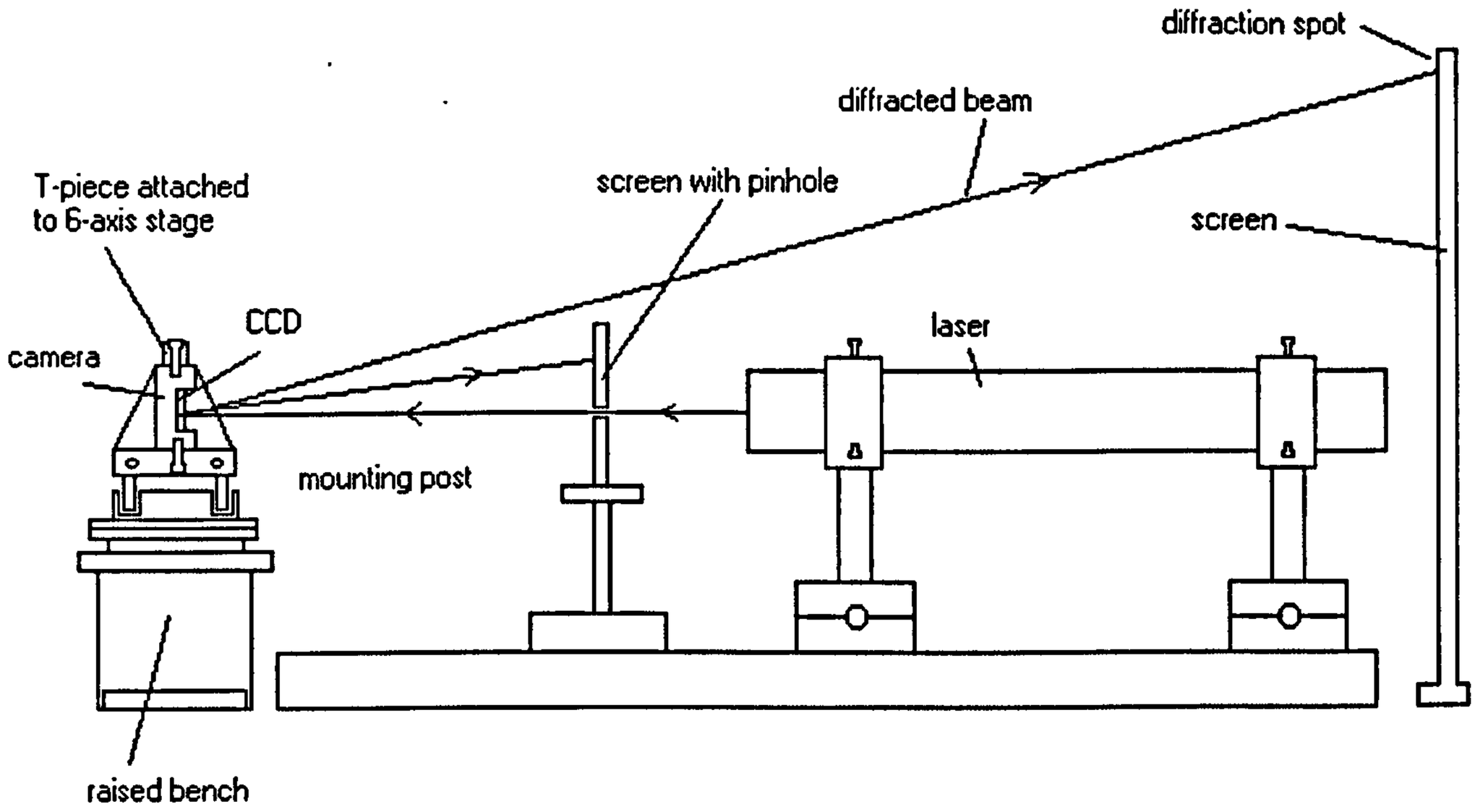

Figure 52 - camera alignment setup, side view

5.3.5 Camera fixing

The mount consists of two sections. One is fixed to the camera, and incorporates three pins which protrude downward into pots in the second section, which is fixed to the camera rail. The camera is translated vertically so that the pins are fully out of the pots in the rail plate, and glue is poured into the pots. The camera is then translated back down into the pots, and the alignment is checked to ensure that the diffracted points still fall on the crosshairs. If they do not, adjustments are made so that they do. The apparatus is then left in position until the glue has set, after which time the six-axis mount's T-piece is removed from the camera. The alignment is checked again, and once alignment is verified, the camera is known to be permanently set with the CCD parallel to the rail it is mounted on.

5.3.6 Second camera alignment

The procedure for aligning the second camera is exactly as for aligning the first, only this time the reference crosshairs are fixed. First, the rail is moved and the alignment stage repositioned so that the horizontally diffracted points from the second camera fall on the crosshairs, at which point it is aligned with precisely the same CCD orientation in roll, pitch, yaw and z-direction as the first camera.

The cameras are partially aligned in the y-direction by using the physical surfaces of the camera bodies. Section 4.1.11 shows that the cameras need only be aligned in the y-direction to an

accuracy of 0.838mm, and the housing is specified to be 62mm high to within 0.20mm[28]. This gives the centre of the housing as 31mm ± 0.10mm from the bottom edge, and the lens is to be mounted relative to the centre of the housing in the C-mount, so the centres of the bottom edges are made an equal height from the bench using callipers.

The camera is glued in an identical way to the first, so that the rail has two glued and aligned cameras adjacent to each other. This concludes the first phase of the stereo camera setup, ensuring that all subsequent alignment errors are purely translational in the x, y, principal distance and principal point shift directions, and not rotational.

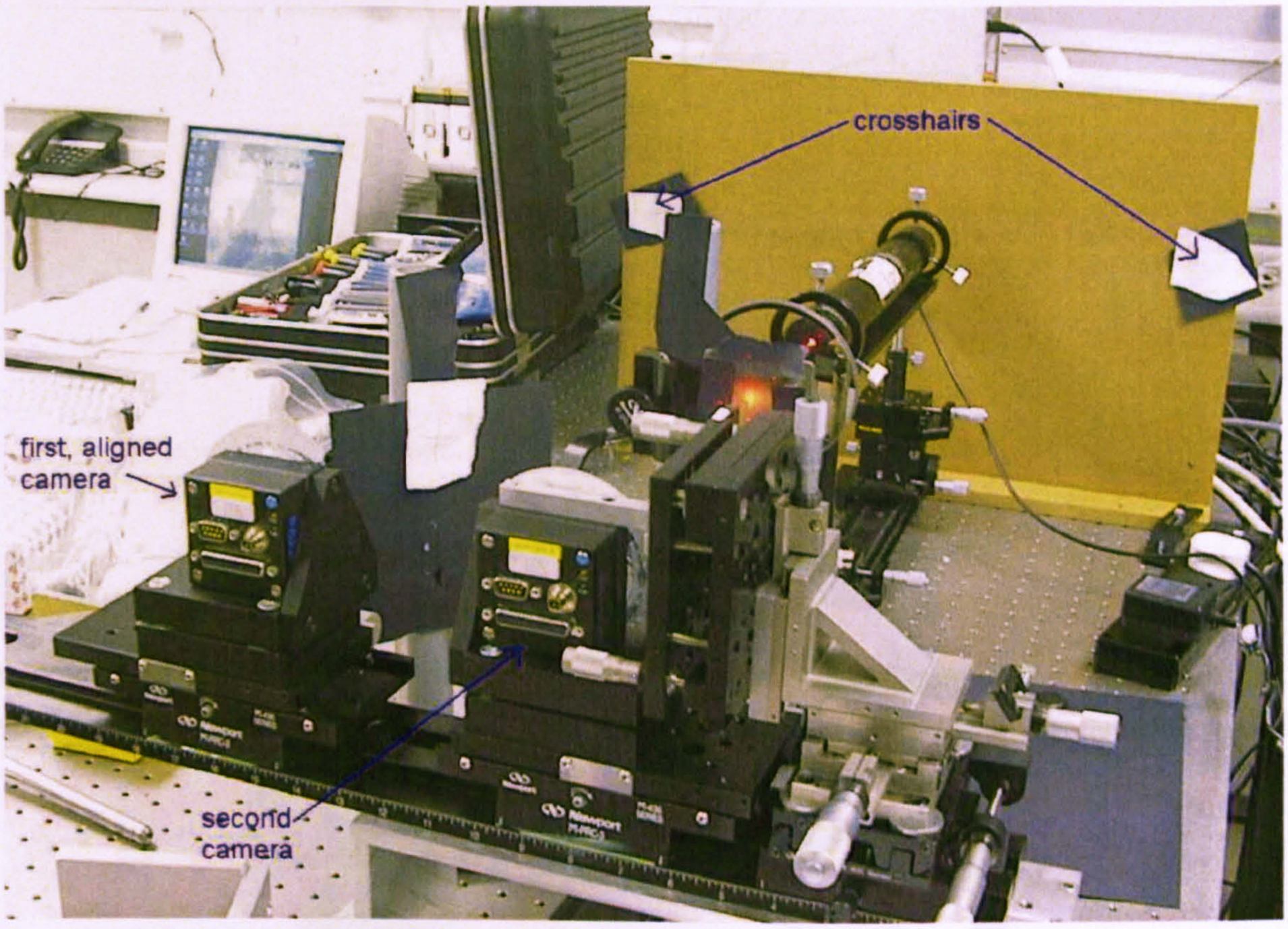

Figure 53 - aligning the second camera. The diffraction spots are not visible in this photograph, as the path of the laser is blocked.

5.3.7 Potential sources of systematic error in the procedure

The lens mount in the Basler A101CP contains a shield to protect the CCD from contamination by dust or other unwanted substances. The shield is part of the lens mount assembly, as shown in Figure 54.

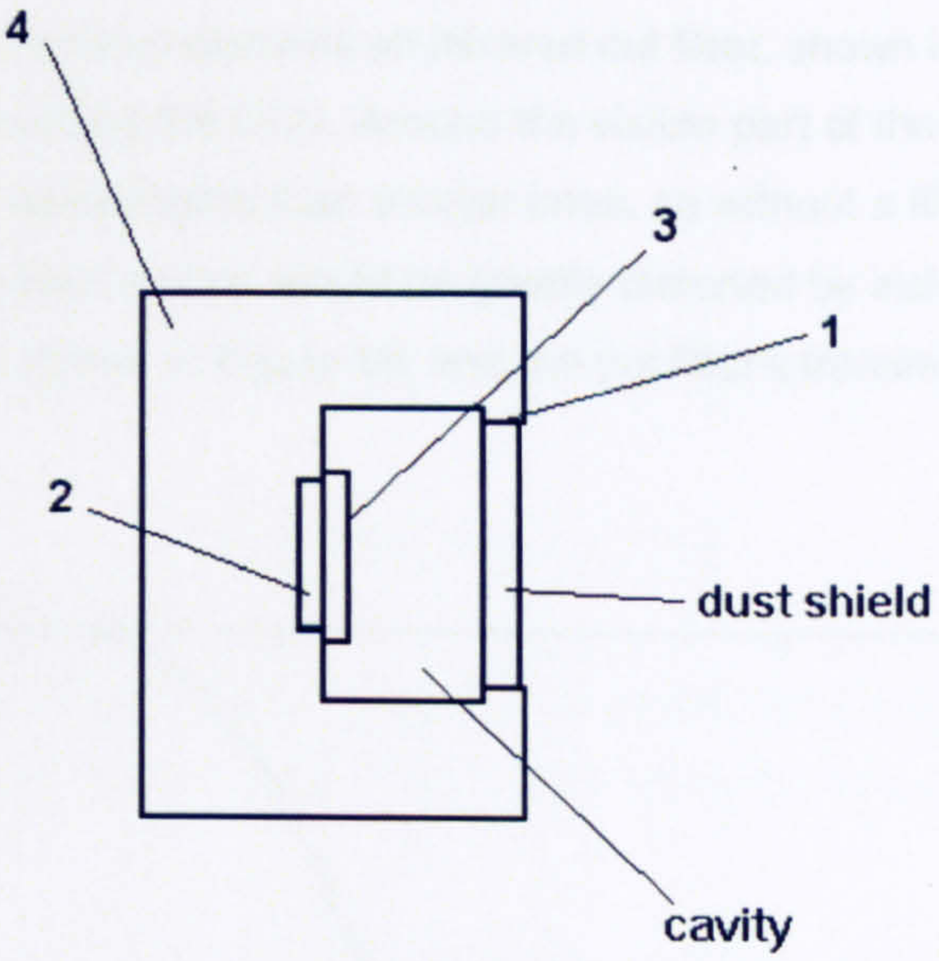

Figure 54 - diagram of the interior of the camera

In Figure 55, the dust shield has been removed. In Figure 54 and Figure 55 the C-mount is labelled 1, the IR cut filter is labelled 2, the CCD is labelled 3 and the camera body is labelled 4.

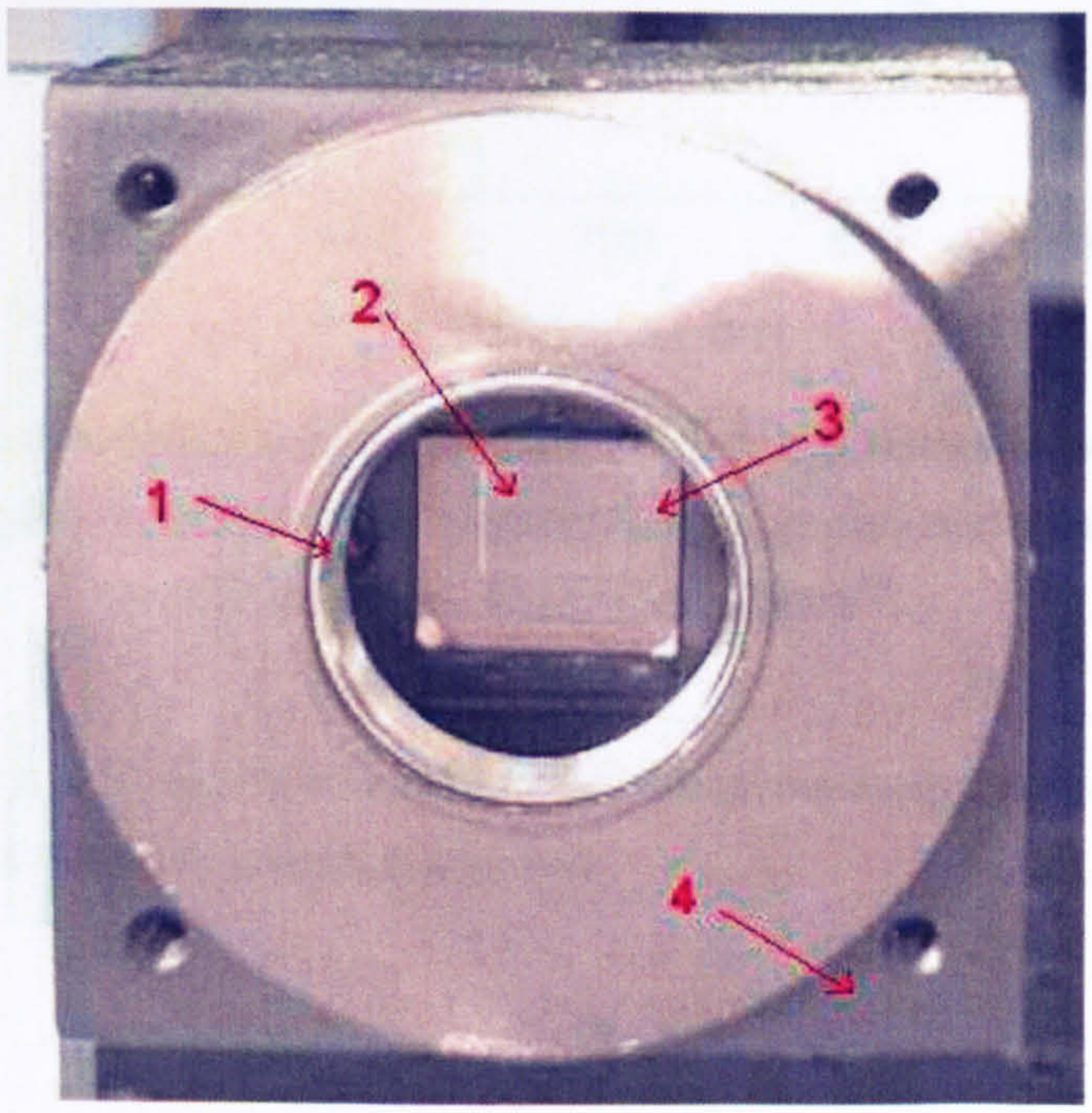

Figure 55 - the interior of the camera

The dust shield is made from glass and is not assured to be perfectly flat or parallel to the CCD, and hence may cause a deviation of the laser beam as it refracts through the glass. This effect can be indirectly seen by observing a second laser spot, close to the zero-order diffraction spot, which has been reflected from the front surface of the glass. This glass shield must be removed in order for the results from the calibration to be genuine. Care must be taken during the camera's lifetime to keep dust away from the sensor.

The Basler A101CP camera also contains an infrared cut filter, shown in Figure 54, which stops infrared radiation from reaching the CCD. Around the visible part of the spectrum, CCDs are more sensitive to longer wavelengths than shorter ones, so without a filter the CCD's red response, and thus the colour output, would be greatly distorted by incident infrared light. The CCD's response curve is shown in Figure 56, and the cut filter's transmission is shown in Figure 57.

Figure 56 - Spectral response curve for a Sony 2/3" interline transfer HAD CCD. Wavelengths longer than about 700nm are infrared and may distort the red response of the camera. Figure from Sony[93].

The filter removes this light, and enables the camera to give a more balanced response to different colour channels in the images it acquires.

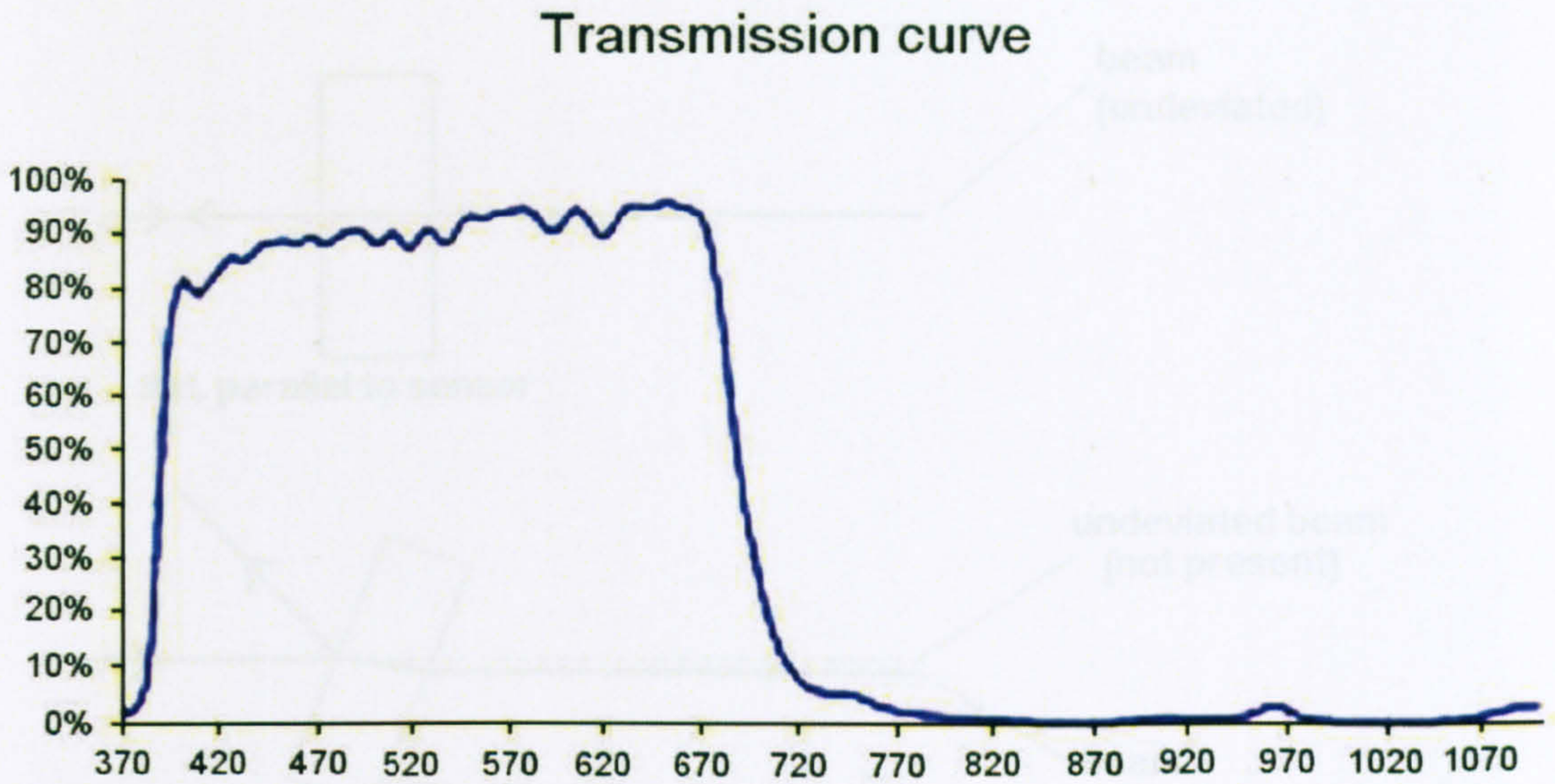

Figure 57 - the transmission of the Schneider B+W 486 filter. Most of the infrared and ultraviolet light is blocked (adapted from documentation[94])

The filter used in the A101CP is a Schneider B+W 486 UV/IR cut filter[94], which is precision manufactured to have plane-parallel surfaces to a quarter of the operating wavelength. This precision ensures that the angular deviation of the laser beam is not significant. The filter could still be mounted not parallel to the CCD, which would not affect the calibration but could show a second reflected spot from the front of the filter, however in practice this was not observed. Effects of a filter not flat or not parallel to the sensor are shown in Figure 58. For these reasons, the IR filter is not considered a problem in this case in the construction of the stereo camera.

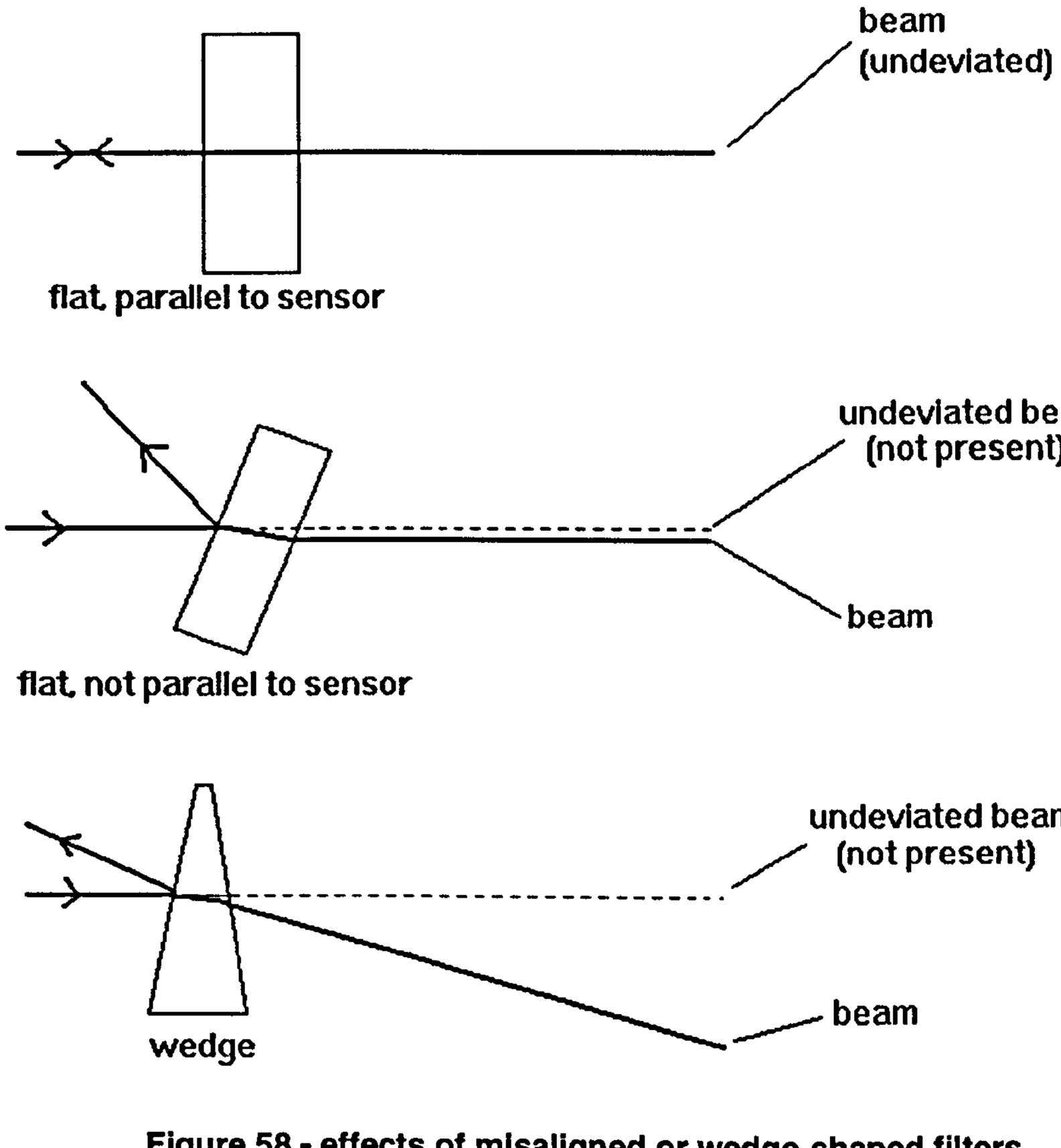

Figure 58 - effects of misaligned or wedge-shaped filters

5.3.8 The quality of the alignment

Testing the alignment of the CCDs is simple, requiring only that the rail is translated so that the beam falls on each CCD in turn. This check showed that the diffraction patterns (and hence the cameras) are aligned as well as can be achieved by eye. The alignment errors are detailed below in section 5.3.9.

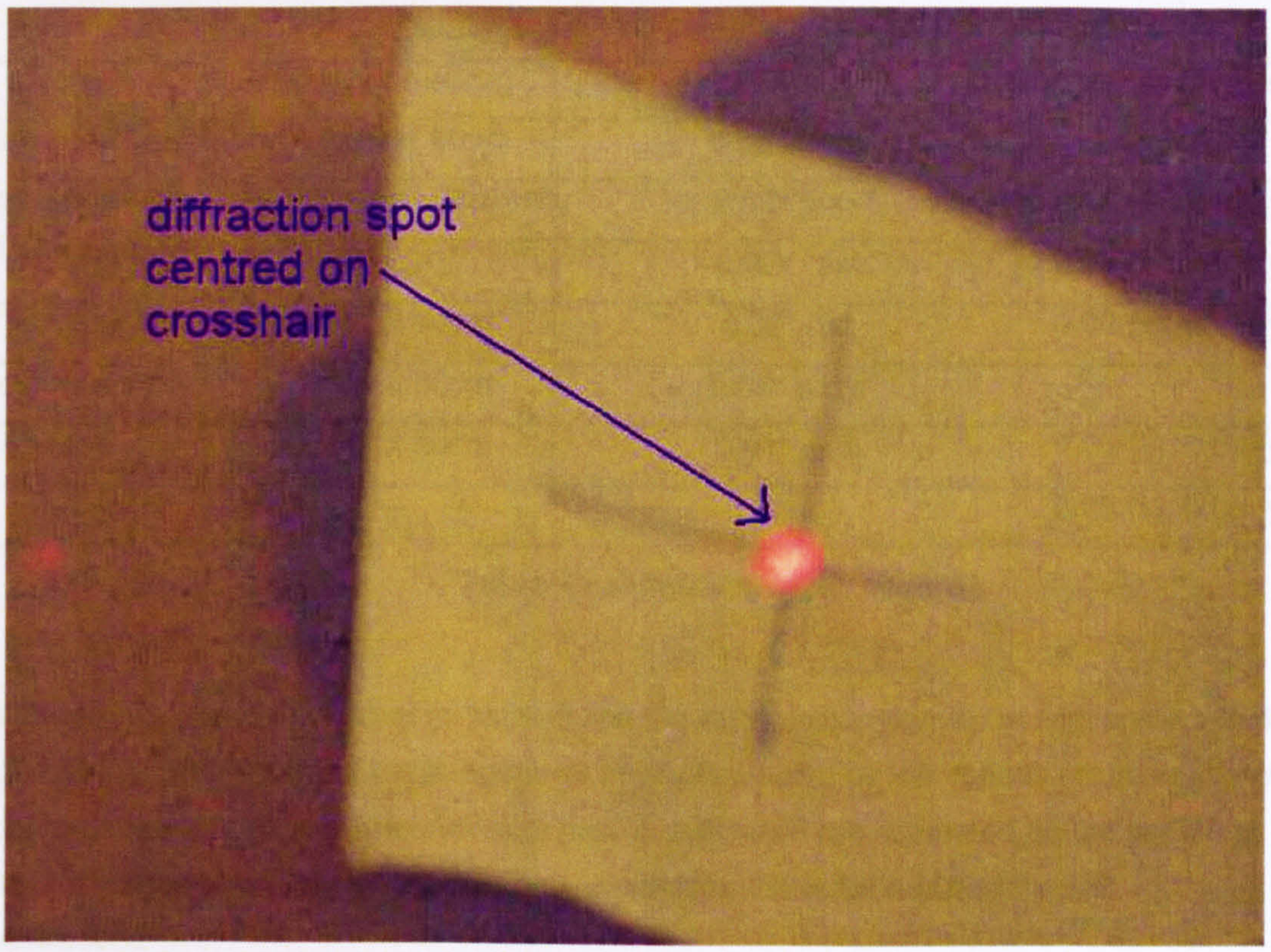

Figure 59 - the right hand spot alignment for the right camera.

5.3.9 The errors in alignment

There are several sources of possible error in the alignment procedure.

Calculations of the slidebar-type errors in Table 2 below are made using the expressions derived in section 4.1.17 to determine the image error in pixels caused by the various errors in aligning the beam, rails and first camera. Spot movements for beam alignment and 1st camera yaw were estimated on the screen at 30cm, and those for the camera roll and pitch were estimated from the bench slide measurements out to 60cm. The total is calculated as the square root of the sum of the squares (RSS) of the individual errors.

Type	Size (spot motion)	Error	Pixel error (axis)
Beam alignment	0.1mm over 30cm	0.02° pitch	-
Beam alignment	0.1mm over 30cm	0.02° yaw	0.02 (x), 0.02 (y)
1st camera rail	-	0.05° yaw	0.06 (x), 0.05 (y)
1st camera roll	0.1mm at 60cm	0.02° roll	0.00 (x), 0.04 (y)
1st camera pitch	0.1mm at 60cm	0.01° pitch	-
1st camera yaw	0.2mm at 30cm	0.04° yaw	0.05 (x), 0.04 (y)
Total	-	-	0.08 (x), 0.08 (y)

Table 2 - slidebar errors

The single-camera errors shown in Table 3 use the expressions derived in sections 4.1.7, 4.1.8, 4.1.9 and 4.1.12 to calculate the image error in pixels caused by the various errors in aligning the second camera. All camera error spot motion estimates are estimated on the screen at 80cm, the translation stage errors are upper limits from Newport's 443 series rail specifications[95].

Type	Size (spot motion)	Error	Pixel error (axis)
2nd camera roll	0.1mm in 90cm	0.02° roll	0.17 (x), 0.21 (y)
2nd camera pitch	0.2mm in 90cm	0.01° pitch	0.03 (x), 0.16 (y)
2nd camera yaw	0.2mm in 90cm	0.01° yaw	0.17 (x), 0.03 (y)
2nd camera Y-axis	-	0.10mm ΔY	0.00 (x), 0.12 (y)
2nd camera Z-axis	0.1mm in 90cm	0.265mm ΔZ	0.17 (x), 0.14 (y)
1st translation stage roll	-	0.01° roll	0.16 (x), 0.13 (y)
1st translation stage yaw	-	0.01° yaw	0.27 (x), 0.05 (y)
2nd translation stage roll	-	0.01° roll	0.16 (x), 0.13 (y)
2nd translation stage yaw	-	0.01° yaw	0.27 (x), 0.05 (y)
Total	-	-	0.53 (x), 0.38 (y)

Table 3 - single camera errors

Combining both sets of alignment errors (using RSS) gives disparities for the stereoscopic camera pair due to errors in alignment, summarised in Table 4:

	Δx	Δy
error/pixels	0.54	0.39

Table 4 - total pixel errors for the alignment procedure

5.4 Conclusions

The laser alignment can be a valuable tool for aligning the first camera to the rail, but given that machine vision cameras and optical rail mounts are generally manufactured to a very high specification, how useful is it for aligning every camera? The answer to this question determines whether each camera must be individually aligned by laser, or whether every camera can simply be mounted from the same 6-axis stage (with the same settings) as the first camera.

Before aligning the second camera with the laser, a measurement of the spot locations compared to the spot locations from the first camera's diffraction pattern was made. This shows the differences in sensor position relative to the mounting between the two cameras. The differences in the spot positions were:

First spot, (+6mm, +8mm)
Second spot, (+7mm, +6mm)

Given the screen distance of 800mm, this indicates a camera misalignment of approximately 0.2° in roll, and 0.5° in pitch and yaw. This error is made up of two components, one due to the positioning of the sensor relative to the mounting, and one due to the error in repeating the six-axis stage's slider orientation on the rail. This error would produce significant image errors of about 5.4 pixels in the x direction and 5.1 pixels in the y direction, and so the laser alignment must be used for each and every camera.

The method has the advantage not only of high accuracy, but of simplicity. The process lends itself to automation, since feedback from two spot detectors can in principle control actuators to adjust the alignment apparatus.

An alternative would be to use cameras where the sensor is aligned relative to the camera's mounting points to within the required accuracy. Then, the cameras could be aligned simply by physically aligning the mountings. However, this would require the sensor alignment to be set using a similar method to the one used here for the whole camera.

5.5 Summary

This chapter has shown how using a laser to align the sensors in a digital stereoscopic camera system can achieve a high level of accuracy, which in the case of the camera system described above produces a vertical image error of only 0.39 pixels. This concludes the first part of the

camera calibration, and chapter 6 describes the second part, that of calibrating the lens systems.

6 Lens calibration

6.1 Method

6.1.1 Introduction

Of the nine errors possible in the simple pinhole error model adopted in section 4.1.2, six are eliminated by the laser alignment – pitch, roll, yaw and x-, y- and z-translations. This leaves only the x-centre, y-centre and principal distance errors to be addressed by a calibration of the lens. These three errors are caused by translations of the lens centre in the X-, Y- and Z-directions.

6.1.2 Translation errors

The x-centre and y-centre translations arise from misalignments of the optical centre of the lens with the centre of the sensor arrays, and produce two-dimensional translations of the images within their own plane. If the misalignments of the two cameras' optical centres are not the same, one image of the stereo pair will appear translated with respect to the other. The error could be eliminated by fixing the optical centre to lie on the same principal axis in each camera, however this is impractical for three reasons. Firstly, locating the optical centre with pixel or better accuracy is difficult[47]. Secondly, lens assemblies and their mountings are not normally rigid. Screw mountings such as the C-mount allow motion between the two threads, and lens systems are generally unstable to being moved, adjusted or knocked. Finally, the camera requires that the lens be able to focus, and the action of focusing the lens involves moving either the whole lens assembly or elements within it, which moves the optical centre and the principal point, notably where the optical axis of the lens is not aligned with the mechanical axis of the focusing system. Eliminating this error in hardware is therefore very difficult. However, cropping an image pair to remove a translation between them is a simple and nondestructive operation, so a software correction is a satisfactory solution. The only drawback to this method is that the final image size will be smaller (though only slighly smaller for a small translation correction) than the maximum available from the image array if areas are cropped out. This could be helped by using an intentionally oversized sensor, or an undersized output image.

In order to correctly crop the image, the amount of cropping must be determined. Automatic cropping requires stereo matching, which is prone to inaccuracy and error[96]. Most stereo

matching schemes are tailored to a particular scene structure[96], and do not perform uniformly over all conceivable scenes. The most fundamental problem facing automatic matching schemes is that of occlusion, as it is not possible to match a feature or area on one image to one which does not exist in the other[97]. Automatic cropping schemes based on scene analysis must also overcome problems including brightness or colour differences between the two images, reflections, repetitive features, featureless regions and other variety in the types of image concerned, while a cropping based on a calibration of the hardware performs measurably and uniformly for any type of scene. Such a calibration-based cropping requires less on-demand computation than scene analysis methods, and can even give a correct cropping value for featureless images. So, any lens calibration should calculate the cropping required for a given image pair.

Because of the instability discussed above, this may not be pixel-accurate, but the calculated cropping can assist software in narrowing down the range of possible matches for a more accurate cropping determination.

To avoid focusing errors as described above, a scheme might be constructed where the lens is held firmly in place, while the sensor is translated back and forth to provide focusing control. Such a scheme is not used here, for two reasons. Firstly, lens focusing mechanisms are available as standard features on lens systems, and are largely self-contained. Detaching the lens from the mounting on the camera body for a basic body translation focusing mechanism would expose the sensor surface (or any filter on top of it) to environmental hazards such as dust or fumes. Secondly, the translation stages available for test quote and show a high tolerance on angular deviation[95], but do not quote tolerances on linear deviation from the translation axis. Body translation focusing experiments with the Newport translation stages[95] show unrepeatable image shifts of several pixels, thought to be due to contamination of the translation stage rails with dust. For these reasons, the conventional focusing method of using the lens system's focusing barrel is used at this time.

6.1.3 Scaling errors

Principal distance errors arise through the optical centres of the two lenses being located at different distances from their sensors. This produces an inequality in the fields of view between the two cameras, and so one image appears scaled with respect to the other. When taking an image, if each camera is focused to give maximum sharpness at the same point, any difference in principal distance is due to a difference between the focal lengths of the two lenses. If the two lenses have significantly different focal lengths, then they cannot both produce sharp images at the same scale unless the aperture is very small. Given an operating aperture for the camera and a limit for the out of focus blur, the required closeness of the focal lengths of the two lenses can be calculated.

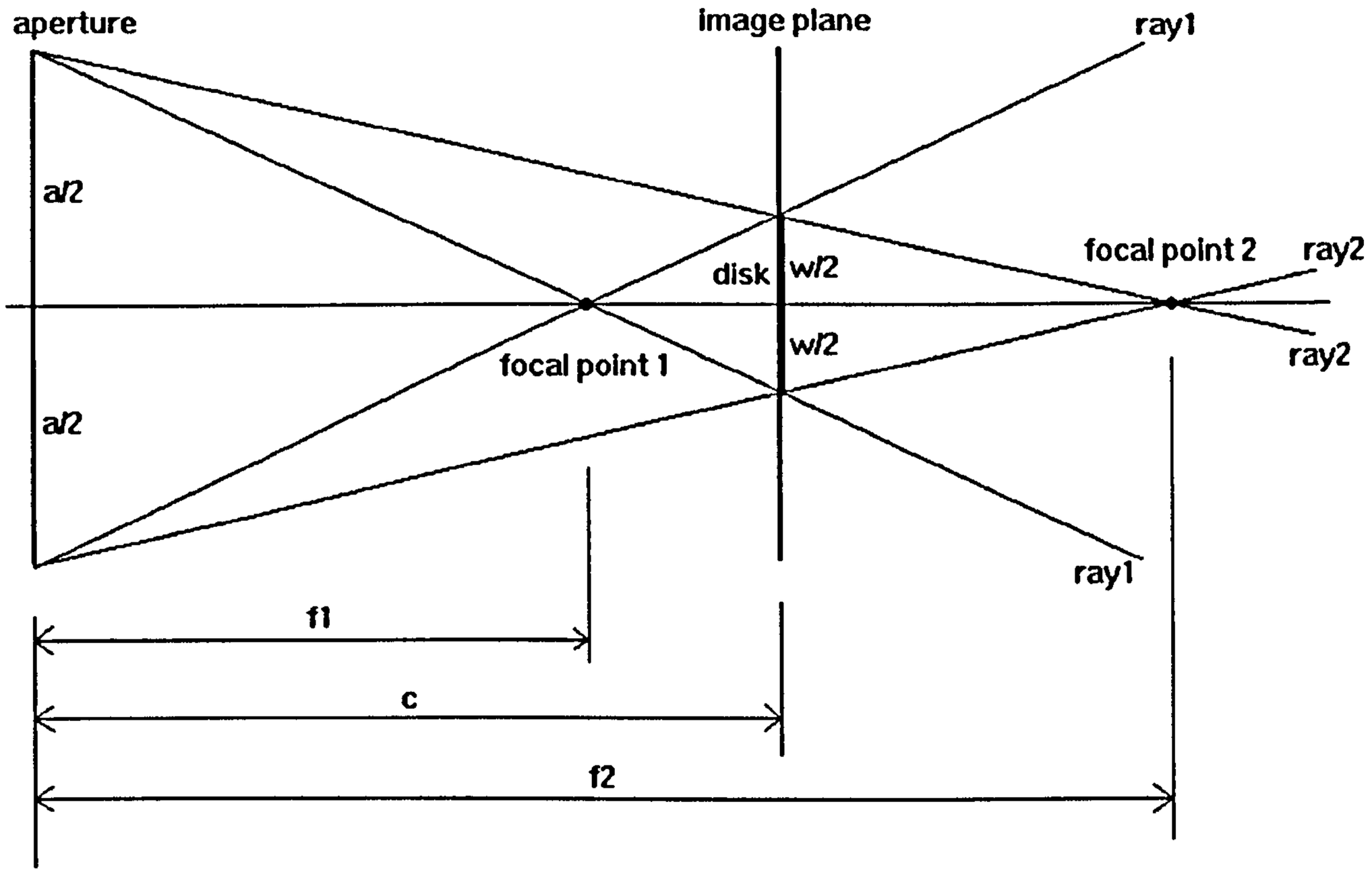

Figure 60 - overlaid ray diagrams for a pair of cameras with different focal lengths

The most flexible arrangement has the image plane behind the focal plane for the camera with the shorter focal length, and the image plane in front of the focal plane for the camera with the longer focal length. Figure 67 shows the arrangement.

The aperture has diameter a, the out of focus disk has diameter w, the focal lengths of the two cameras are at distances f_1 and f_2 from the optical centre, and the image plane is at a distance c such that the disk has the smallest maximum size possible in either image.

By similar triangles:

$$\frac{\left(\frac{a}{2}\right)}{f_1} = \frac{\left(\frac{w}{2}\right)}{c - f_1}$$

$$\frac{\left(\frac{a}{2}\right)}{f_2} = \frac{\left(\frac{w}{2}\right)}{f_2 - c}$$

Equation 42

So:

$$c = f_1\left(1+\frac{w}{a}\right) = f_2\left(1-\frac{w}{a}\right)$$

$$\frac{f_2}{f_1} = \frac{1+\frac{w}{a}}{1-\frac{w}{a}}$$

Equation 43

For the Basler A113C camera with the Schneider 8mm lens at f/2.8, a maximum blur of 1 pixel (6.7μm), w = 1 pixel and a = 426.44 pixels. This gives the ratio of f_2 to f_1 as 1.0047. So the focal lengths must be within 0.47% of each other in order to give images of the same scale with a one pixel focal sharpness.

In practice, it is difficult to achieve the flexibility of the above arrangement, because it is difficult to measure a one-pixel blur on an object every time an image is taken. An easier arrangement has one camera being focused as well as possible on an object, while the other camera is slaved to produce an image of the same scale. This effectively makes one of the focal lengths equal to c in the above derivation, and reduces the maximum difference in focal length of the two lenses by half to 0.24%.

A lens calibration should therefore be able to test whether or not the focal lengths of a pair of lenses are within this tolerance, since a pair of lenses outside the tolerance are unsuitable for a camera which is to take images that do not require scaling in post-processing. Because the camera will be required to focus on different objects at different times, and the second camera is to be slaved to the first in regard to image scale, the calibration will need to provide enough information to do this as well.

6.1.4 Requirements

The calibration needs to allow a pair of cameras to be matched in image scale, and to calculate the translation between the images on the basis of the lens focusing. Since the lens is responsible for forming the image, it makes sense to use an imaging method for the calibration. The actual calibration must measure the x,y position of a target as well as the scale, at a range of lens focus points which is sufficient to describe the scale and position at any lens focus point to be used. In order to image a target over a wide range of lens focus positions, the lens needs to be operated at a very small aperture to enable the image to be in reasonable focus throughout the calibration. The target and its analysis must also be robust to some degree of

out of focus blurring, since this will occur even at a lens's typical minimum aperture of around f/22.

6.1.5 Basic features of the calibration

Though many camera calibration methods exist[46], the existing alignment of the camera sensors detailed in 5 simplifies the requirements of this camera's calibration, as only principal point and distance must be calibrated, and these only need to be calibrated in each camera relative to the other. This simplification allows for the used of fixed cameras and target, and for the target to be planar, which makes the calibration process simpler and more portable than less specific calibrations.

Calibration involves two important steps. The first is to identify features or areas in images taken with the cameras, and to determine a set of image coordinates which characterise their positions. The second step is to use these coordinates to calculate the state of the cameras, in this case in terms of principal point and distance.

Since the calibration images will not typically be in sharp focus, identification and measurement of edges or textures is not advisable. Out of focus points bear a close resemblance to stellar images in astronomical photographs, and there are many existing tests of methods for locating such points[98,99,100]. The two most successful methods are the centroid and the Gaussian fit[98]. Tests have shown that the centroid method often tends to perform more accurately[100], especially when the points being measured are large[99] (such large points also being more accurately locateable than small points[99]). Under poor contrast conditions, where the background level of the images is high, the most successful point location method is that of the modified moment analysis[98], an adaptation of the centroid method where only pixels above a background threshold level contribute to the centroid. Since poor contrast between target point and background is to be produced by calibrating at low aperture and high gain (the cameras have a 130ms upper limit on exposure time[28]), large target points should be used along with a thresholded centroid method of point location.

Once the calibration point coordinates have been extracted, they must be converted into a characteristic position and scale for each image. Existing models[101] contain coupled[47] parameters which if not fixed manually would cause inaccuracy in the determination of position and scale, and for simplicity a simpler model is adopted. The position is characterised by the centre of mass of the target points, and the scale is characterised by the mean distance between the points.

Images of a planar target are taken at a series of measured focal positions with both cameras, after which the same set of points is extracted from each image and used to determine the

translation and scale change between any pair of the images. This data is then used to link the focus position of a lens to the scale of the image produced, so that a pair of cameras can be matched, provided the lens focal lengths are found to be similar. One camera is set for optimum focus on the scene to the photographed, and the other camera's focus is calculated and set such that the scales of the images produced are equal. The images are taken, and the calibration data is used to provide an estimate of the translation between the two images for cropping purposes.

During the calibration, the assumption is made that the relative parameters of the two cameras' optical systems do not change as the aperture size changes, where the apertures of the two cameras are the same as each other. Maximum aperture (approximately f/1.4) is required to accurately identify the best focus position of each lens (as in 6.3.1), while minimum aperture (approximately f/22) is required to keep the calibration target in acceptable focus at all necessary focus positions (as in 6.1.6). The aperture setting for taking stereoscopic images will be dependent on the scene brightness and restrictions on exposure time (as in 4.1.18).

Subsequent sections describe the calibration procedure in detail, and 11.2 and 11.3 provide a simple step-by-step summary.

6.1.6 The calibration target

The target can be relatively simple, because it only needs to contain measurable points in a single plane parallel to the image planes of the cameras. If zero lens distortion is assumed, the location and distribution of the points is unimportant, however this would be a stronger assumption than has been made up to this point. With distortion (other lens aberrations have comparatively minor effects on overall image geometry), it is important that the points are distributed as evenly as possible over the image so that the distortion does not bias the calibration in any particular direction. The points should fall comfortably within the field of view of both cameras for all focus positions, so that every point extracted can be tracked across all calibration images. Their extraction must be resistant to small amounts of blurring, and the target must be bright enough that the cameras can produce usably bright images at the small apertures (f/22 on the Schneider 8mm lenses) necessary to fulfil the depth of focus requirements.

The simplest way to extract features from an image is by using a simple threshold to pick out light or dark features. If, as is usual, more than one pixel per feature falls above the threshold, the pixels must be grouped by feature and a point extracted from each group. Using Gaussian-blurred spots is appropriate for the calibration, for several reasons. The use of an intentionally blurred target means that its analysis will necessarily be able to deal with the blur introduced by focusing the lens. The Gaussian blur causes the centre of each spot to be its darkest point,

giving a simple way to estimate the centre before centroiding. This estimate can be used as a check to ensure the analysis is working properly, and also as the centre for the centroid operation. Each spot is radially symmetric, so the calibration target will have the same properties however the features are orientated in the target plane.

The blurred spots were drawn in Paint Shop Pro[102], and arranged in a grid pattern as shown in Figure 61:

Figure 61 - pattern of blurred spots used for calibration target

The pattern was laser printed on to transparencies, and pairs of transparencies were overlaid and taped together to increase the darkness of the spots beyond the printer's range. Four sets of transparencies were mounted on a large A0 light box, with an additional spot roughly in the centre of the arrangement (to allow a simple estimate of overall translation), as shown in Figure 62:

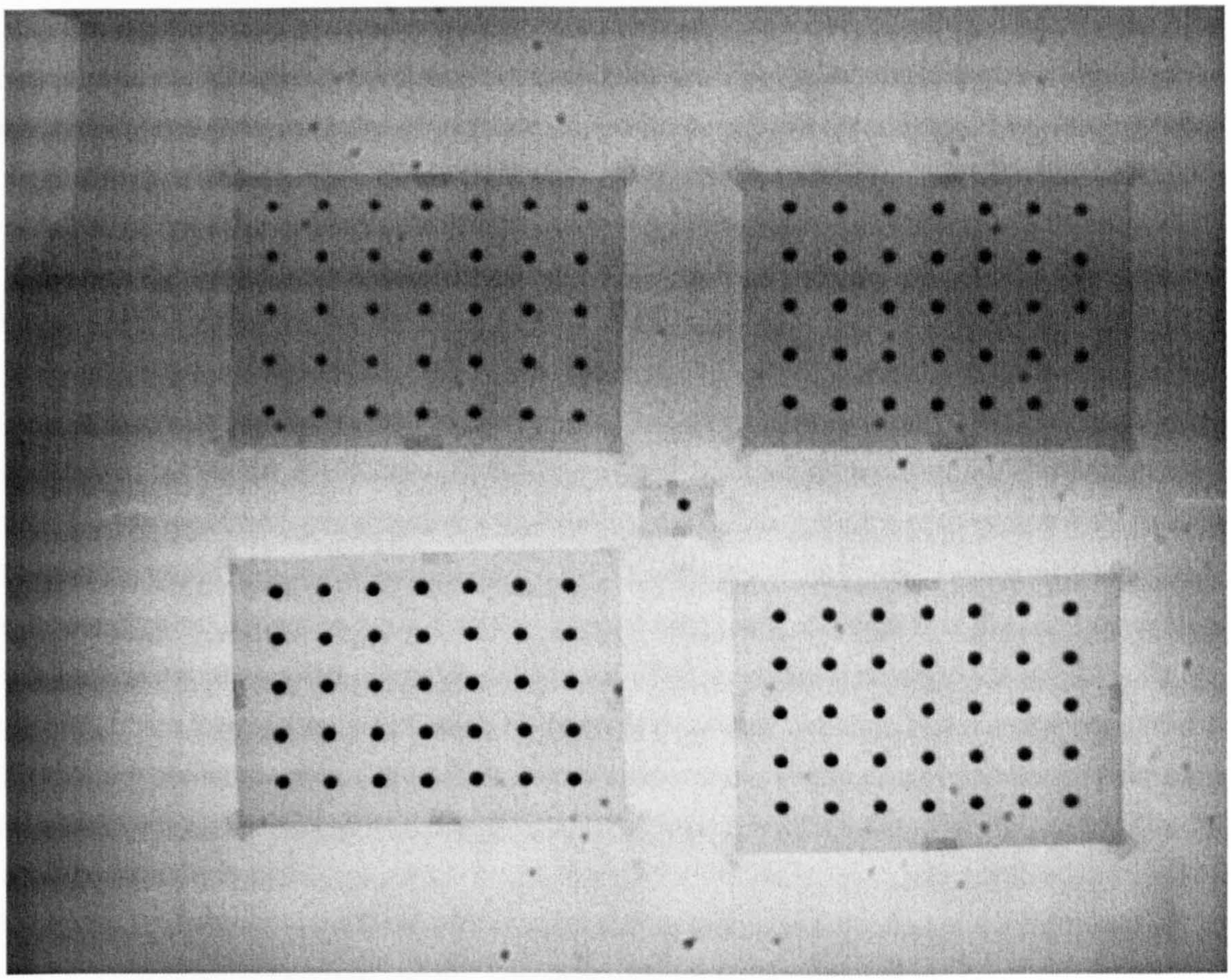

Figure 62 - the target as seen by one of the cameras in a stereo pair

The light box provided a bright background against which the dark spots strongly contrasted. The brightness produced was sufficient to form a clear image of the target at the lenses' minimum apertures at the cameras' maximum exposure time of 130ms, without needing to use high gain settings that would have introduced high levels of noise into the images.

The target was aligned parallel (to within about 1mm) with the camera rail, at a distance of 90cm.

6.1.7 Image acquisition

Images are taken at a sequence of measurable and repeatable lens focus positions for both cameras, at minimum aperture (approximately f/22). The focus positions are measured using the orientation of the rings used to manually focus each lens, and images are taken at a number of regular intervals sufficient to describe the behaviour of each lens. With the lenses used, backlash caused by slack in the barrels threads was not found to be evident, however as a general precaution all measurement runs were performed in the same order from distant to close focus positions.

Each image is interpolated from the initial colour-masked image to a full greyscale image for analysis.

6.1.8 The point extraction method

The method for analysing the images contains a number of simple steps. First, the image is thresholded to pick out the points, which appear dark on the light background. An initial guess is made for the threshold, and the image is scanned row by row. Once a pixel darker than the threshold is found, a point is recognised if it is a specified distance away from every other recognised point. If the new point is close to an existing point, the darker of the two points is assigned as the recognised point. The total number of recognised points in the image is counted, and if the number is not equal to the actual number of points in the image (which is fixed), the threshold is adjusted and the process repeated until the proper number of points are detected.

Each point is then centroided, with every pixel darker than some fraction of the threshold counting towards the centroid weight. The centroids form the output of the point extraction process. Each individual centroid has an associated random error which will vary with the exact conditions and method used. The error will contain components due to inaccuracies in positioning the lens focus ring, shifts in the position of the optics, image noise and the limitations of the detector. For the method and conditions used in section 6.3, the total of these errors is estimated to be 0.36 pixels in the x-direction, and 0.32 pixels in the y-direction, per pixel. This was calculated as the root mean square deviation in the x and y directions, from the mean position across 5 image sets, for 141 points imaged through two cameras. The individual point errors are shown in Figure 63 below:

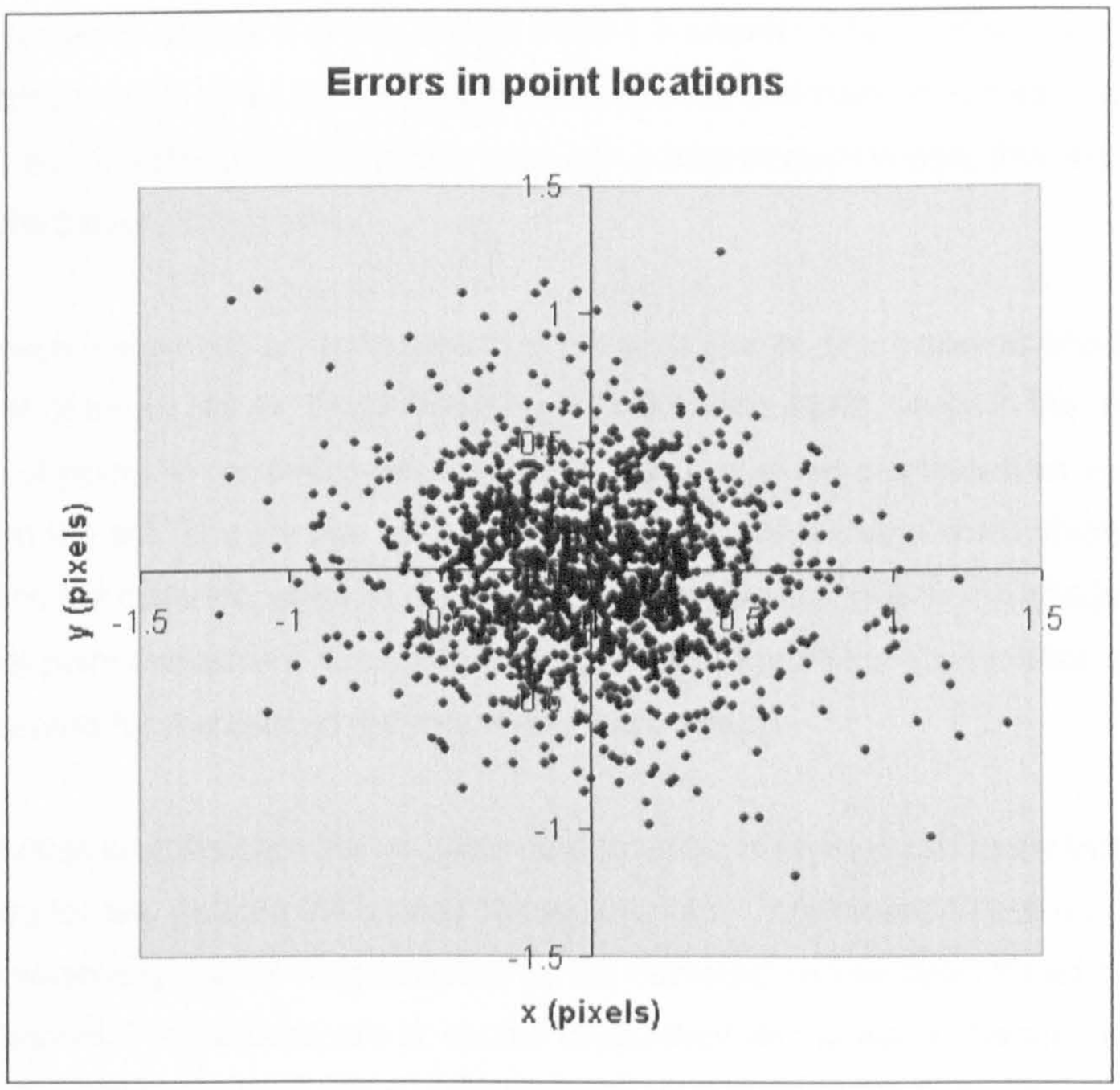

Figure 63 - errors in point locations for 1410 points

With each calibration image containing 141 points, each with a random error as described, the calculated x and y translation of each image is subject to an average error of 0.03 pixels in both the x and y directions.

The above algorithm contains several parameters, including the assumed minimum point separation, the initial threshold estimate, the centroid width and the fraction of the threshold to use for centroiding. These are set manually depending on the properties of the calibration images (such as scale, brightness, contrast) to enable all of the points to be properly detected. No manual intervention is necessary in terms of matching or identifying individual points.

6.1.9 The image change calculation method

The point extraction method produces a set of points which should be the same target points across the image sequence. Each point in the second series is matched to a point in the first, using the closest point in the first. This uses an assumption that the motion of the image points between images is smaller than the distance between image points in a static image, which is valid for a focusing calibration such as this. If any two points in either image set compete for the same match, the closest match is chosen. The process is repeated matching every nth point set with the (n-1)th point set until the whole sequence of image point sets have been processed. Any matches with a distance between matched points greater than a specified value are

rejected as incorrect matches. If in any image a point is unable to be matched, it is erased from the entire matched set, in other words only points which can be tracked across every image are kept. In a data set of 5 images per camera, with 141 points in each image, this algorithm correctly matched every single point.

At this point, each image has an associated set of image points which correspond to the centres of the same set of target points. Each image has a scale calculated, which is the mean distance between pairs of points in the set. Each image also has a position calculated as the average point position in the set. The change in scale corresponds to the image scale change due to changes in principal distance, and the change in position corresponds to the image translation due to principal point movement. These changes produce a profile of the camera, and the process is repeated for the second camera in the stereo pair.

The above process enables the relative scale and position of images produced by the cameras to be calculated for any point in the lenses' focus positions. This means that any the scale and translation between any pair of images taken by the cameras can be determined if the focus positions are known. Translations are of course dependent on the scale change, since the centre of the zoom (the principal point) is undetermined, but since the calibration will be used to match the scales of the two cameras' images, a more complex analysis is unnecessary.

6.1.10 The effect of lens distortion on parameter estimates

Since lenses are generally subject to lens distortion, its effects on the calibration must be estimated. This is performed through modelling the effects of distortion on the calibration target, and the parameters derived from an image of it, and comparing the results with those from a wider and more even point distribution.

The calibration target is modelled as four grids of points, with one additional central point, as shown in Figure 64 below:

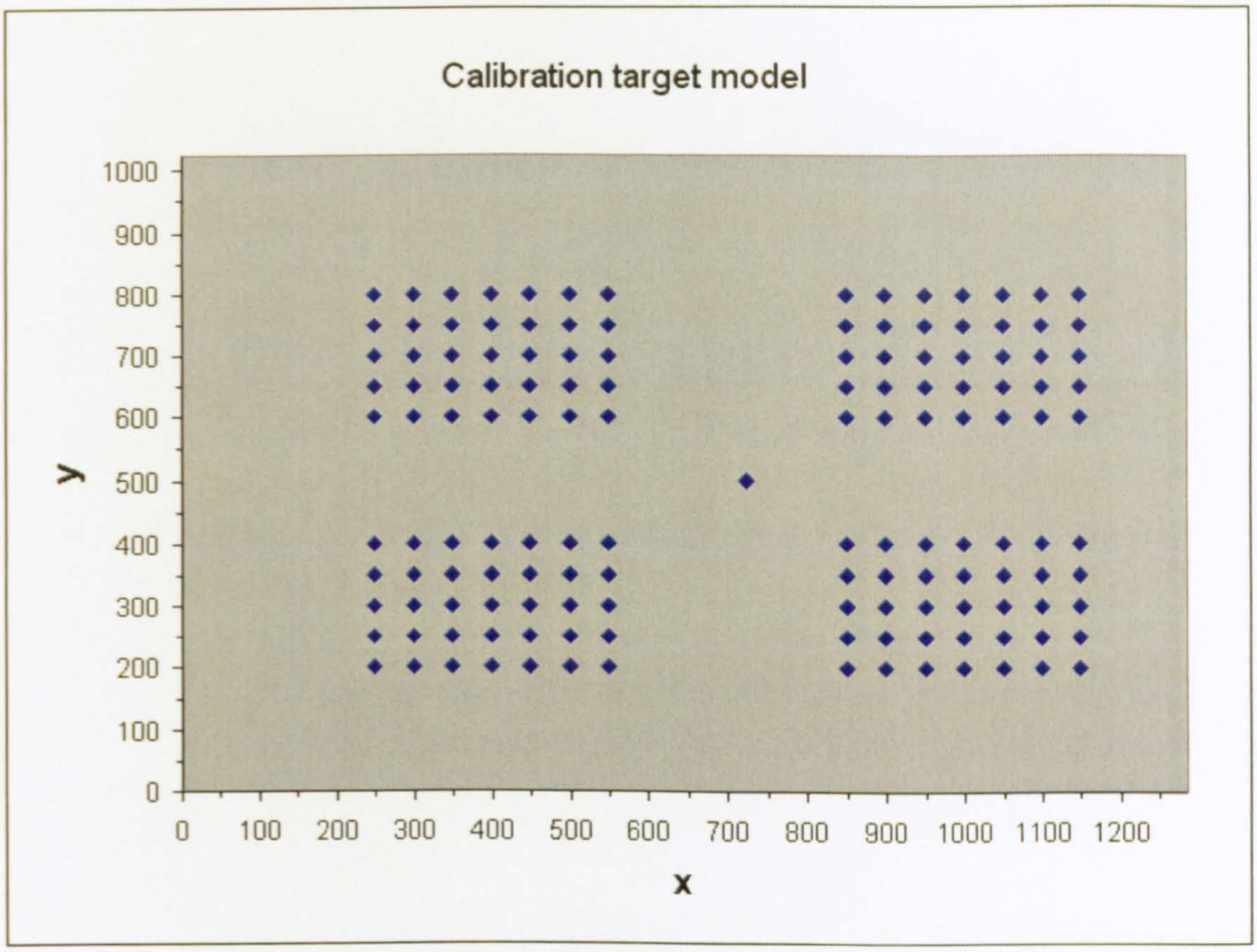

Figure 64 - modelled point distribution of the calibration target

This is an approximation of the calibration pattern seen from a left camera position.

The wide distribution is modelled as 504 points distributed more finely and more evenly, as shown in Figure 65 below:

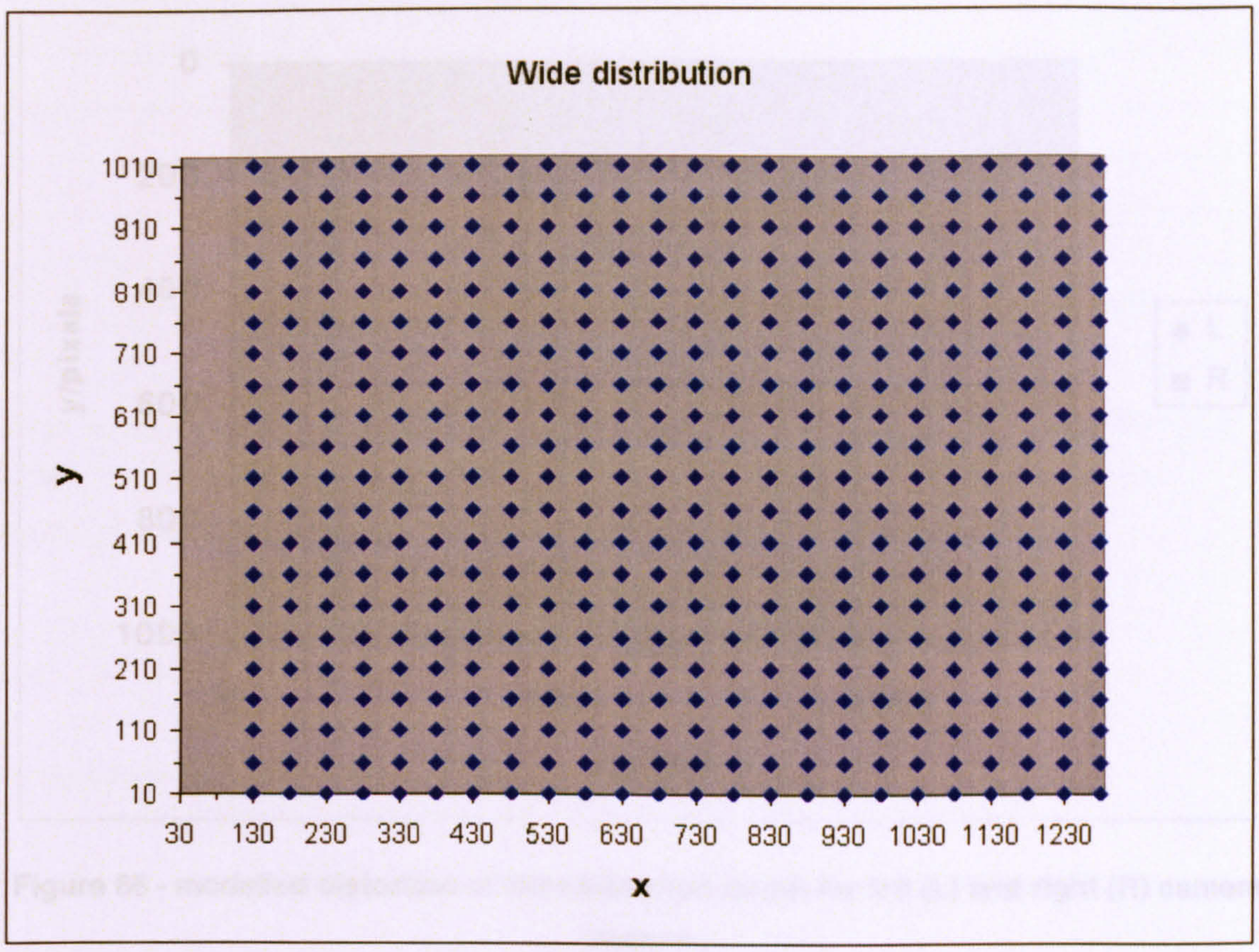

Figure 65 - modelled wide calibration target point distribution

Again, this is modelled for the left camera position.

For the purposes of the model, the parameters show in Table 5 below are assumed:

Parameter	Value
Left to right view x-translation	-130 pixels
Left to right view y-translation	-10 pixels
Radial distortion coefficient, k	4.5×10^{-8} pixels^{-2}
Radial distortion, percentage	~3%
Sensor format	2/3"
Sensor resolution	1284x1024 pixels
Lens focal length	8mm
Modelled distortion centre coordinates	(642,512)

Table 5 - parameters for lens distortion modelling

These values are close approximations of the actual parameters of the stereoscopic camera.

Distorting the modelled calibration target produces the point set shown below in Figure 66:

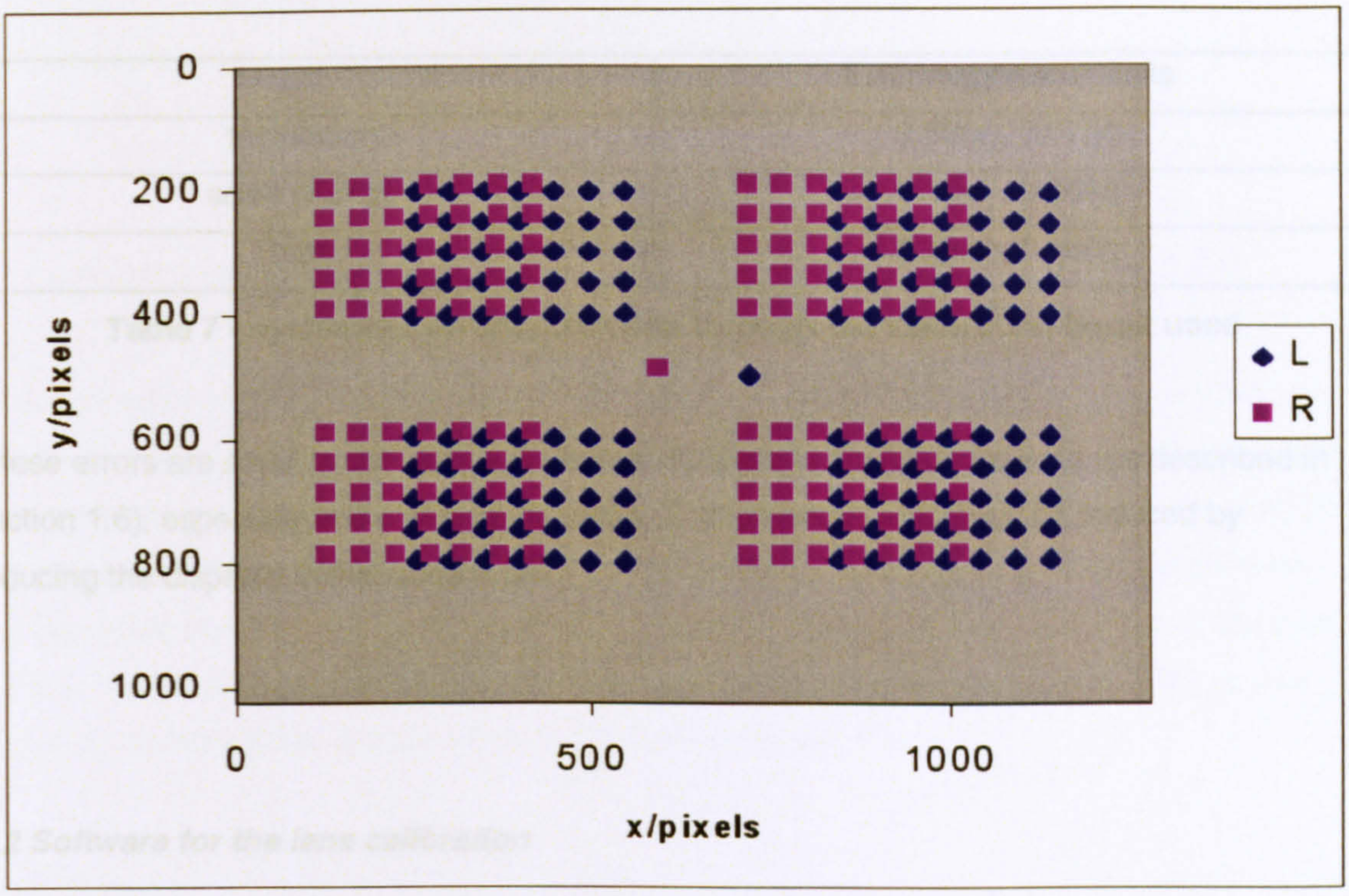

Figure 66 - modelled distortion of the calibration target for left (L) and right (R) camera images

The parameters extracted for the point distributions are shown below in Table 6:

	Undistorted	Target (distorted)	Wide set (distorted)
dx	-130.00	-127.97	-127.34
dy	-10.00	-9.89	-9.82
scale	1.00000	0.99978	0.99994

Table 6 - extracted parameters for modelled point distributions

The wide set is interpreted as an approximately ideal calibration target, and the target set is an approximation of the target actually used. The difference between the target and wide point sets gives an estimate of the errors caused by using a non-ideal calibration target, and these errors are summarised in Table 7 below. All errors are expressed in pixel values, assuming a sensor size of 1284x1024. Errors are treated as systematic, as the target image does not move from one calibration image to the next by an amount comparable to the disparities between the cameras, and are totalled accordingly.

	Error magnitude/pixels
translation	0.63(x), 0.07(y)
scale change	0.10(x), 0.08(y)
total	0.73(x), 0.15(y)

Table 7 - systematic errors produced through the calibration target used

These errors are small in comparison to human tolerances for false disparity (as described in section 1.6), especially in the vertical direction. The horizontal error may be reduced by reducing the disparity in the camera pair.

6.2 Software for the lens calibration

6.2.1 Packages used

The software to implement the method in section 6.1 was mostly written in Microsoft Visual C++, while other parts of the method used commercial software. The sequence in which the various programs were applied is detailed below.

First, the images were captured using the self-written camera GUI. This produced images in the Windows bitmap format[103], which were saved to disk and converted to portable greymap[104] format using Paint Shop Pro[102]. The portable greymap format is a simple ASCII list format, where (after a brief header section) the files consist of a series of decimal greyscale values for each pixel. These files were colour mask-interpolated by a self-written interpolation utility and saved as more portable greymaps. The interpolated greymaps were the input for the self-written point extraction and analysis program, which produced one text file per run containing the positions (the centre of mass of the calibration target points) and scales (mean target point to point distances) of each image. These text files were then pasted into Excel[105], and the data could from there be characterised by a linear fit (in the case of scale changes, which should behave linearly with focus ring rotations of the lenses) or used to look up the translation (which is not so suitable for fitting a function to) for a given lens position.

6.3 Lens calibration

6.3.1 Check for closeness of focal length

As described in section 6.1.3, the lenses are required to have the same focal length to within 0.24%. To check that the lenses used meet this requirement, an image is taken with each lens at its sharpest focus (using high image magnification to help estimate this position by eye), and the scales are calculated and compared. The Schneider 8mm lens pair used in the cameras met the closeness criterion (from section 6.1.3), producing a scale ratio at best focus which was 0.9982. The focal lengths of the lens are therefore within 0.18% of each other, which is enough to give acceptably sharp images when the image scales are matched.

6.3.2 Interpretation of graphs

The remainder of this section contains several graphs which indicate the results of the lens calibrations performed. In those graphs that show the behaviour of a single quantity as the lens is focused, the x-axis shows the position of the lens focus barrel in degrees. The lower end of the scale corresponds to lower principal distance and focus on distant objects, while the higher end of the scale corresponds to higher principal distance and focus on close objects.

6.3.3 Results – left lens

The left lens was calibrated over a range of focus barrel rotational positions of 140° (corresponding to focusing from beyond infinity to within 10cm), with a resolution of 2°, repeated four times to give five full sets of positions and scales. All the figures show changes as a function of the focus barrel position. Figure 67 and Figure 68 show the translation of the target image points' centre of mass as a function of lens focus, and Figure 69 shows the changing scale of the target image, also as a function of the focus position of the lens.

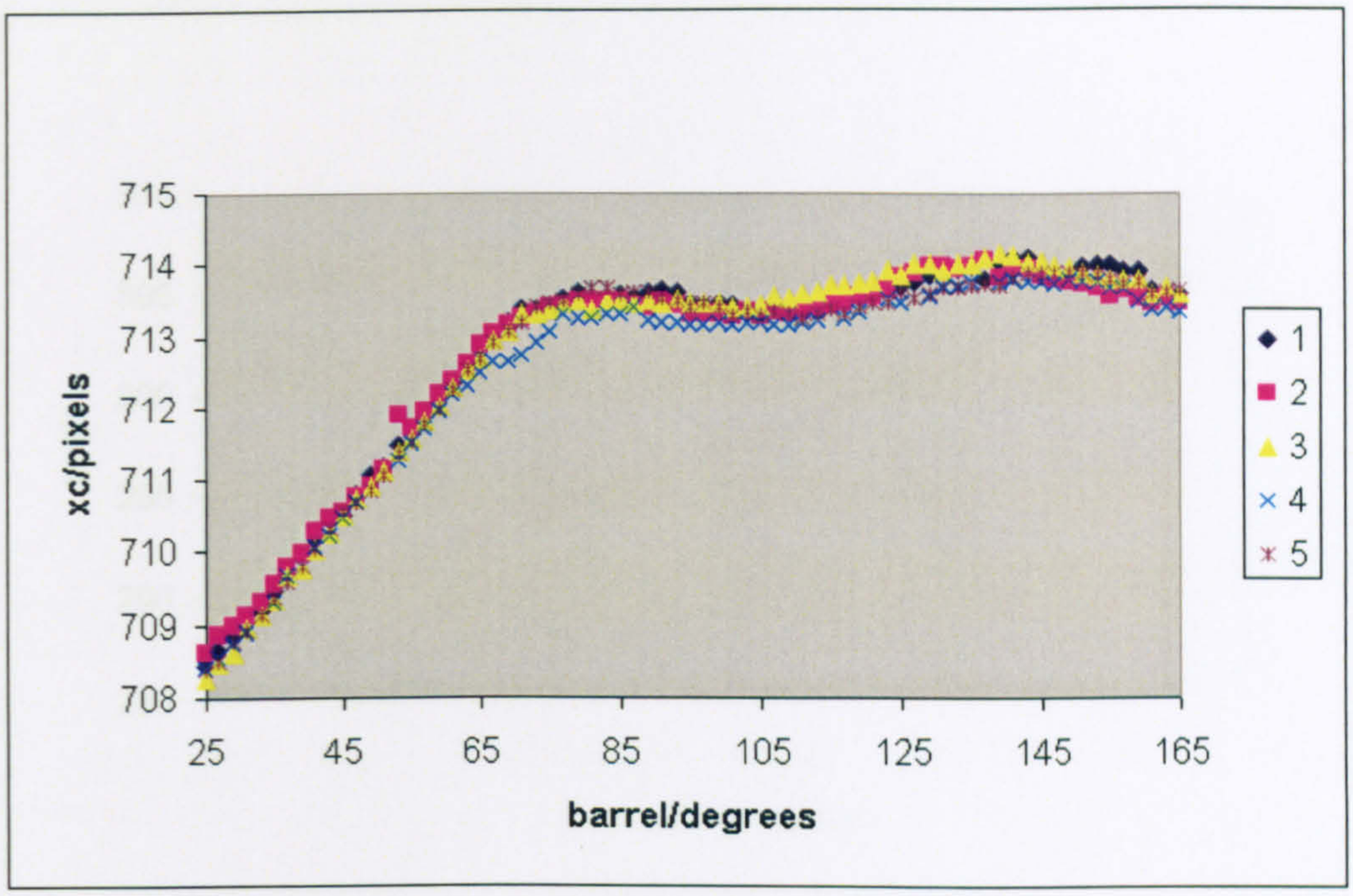

Figure 67 - the change in target centre x-position xc with varying lens focus, left camera

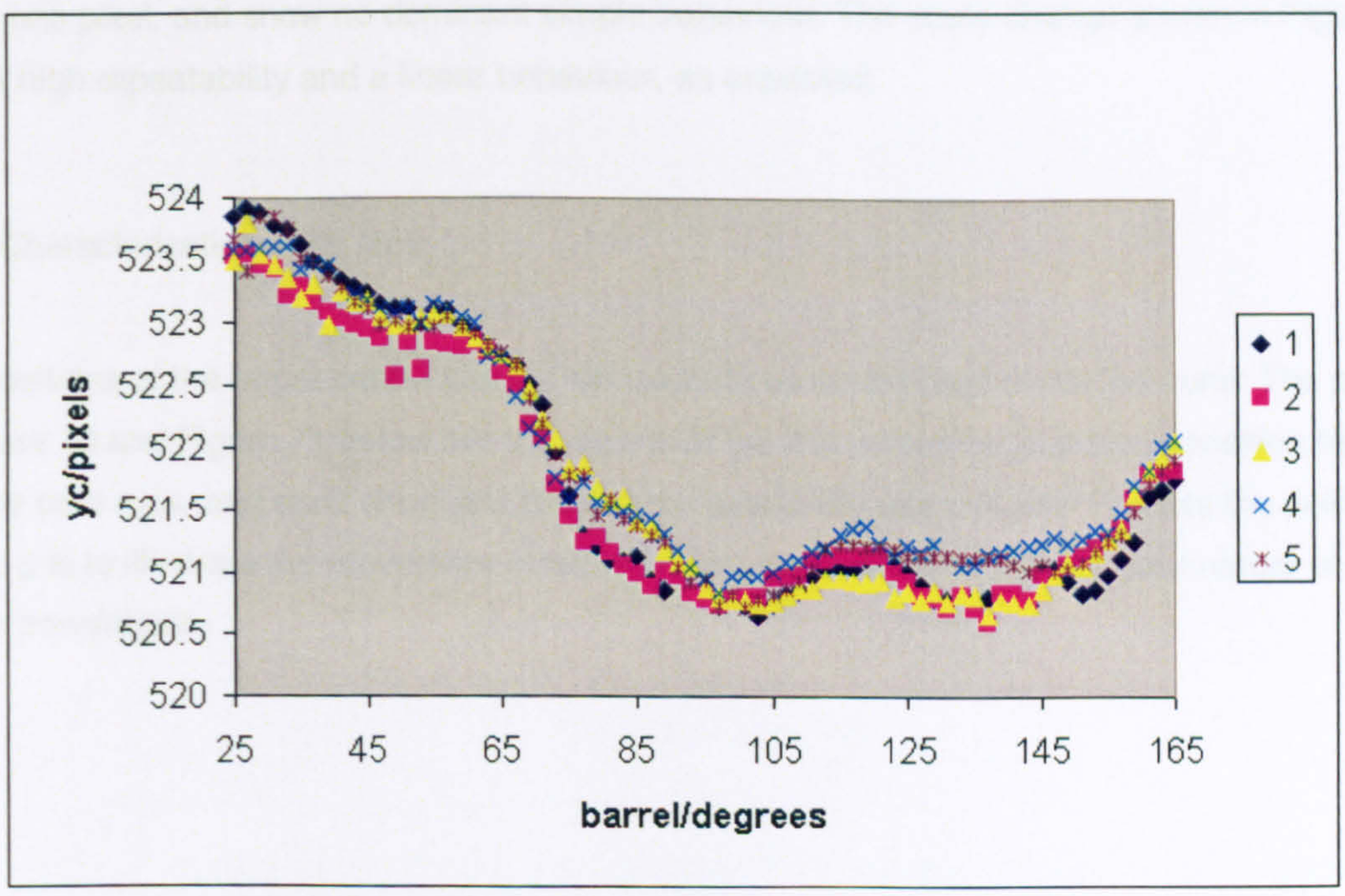

Figure 68 - the change in target centre y-position yc with varying lens focus, left camera

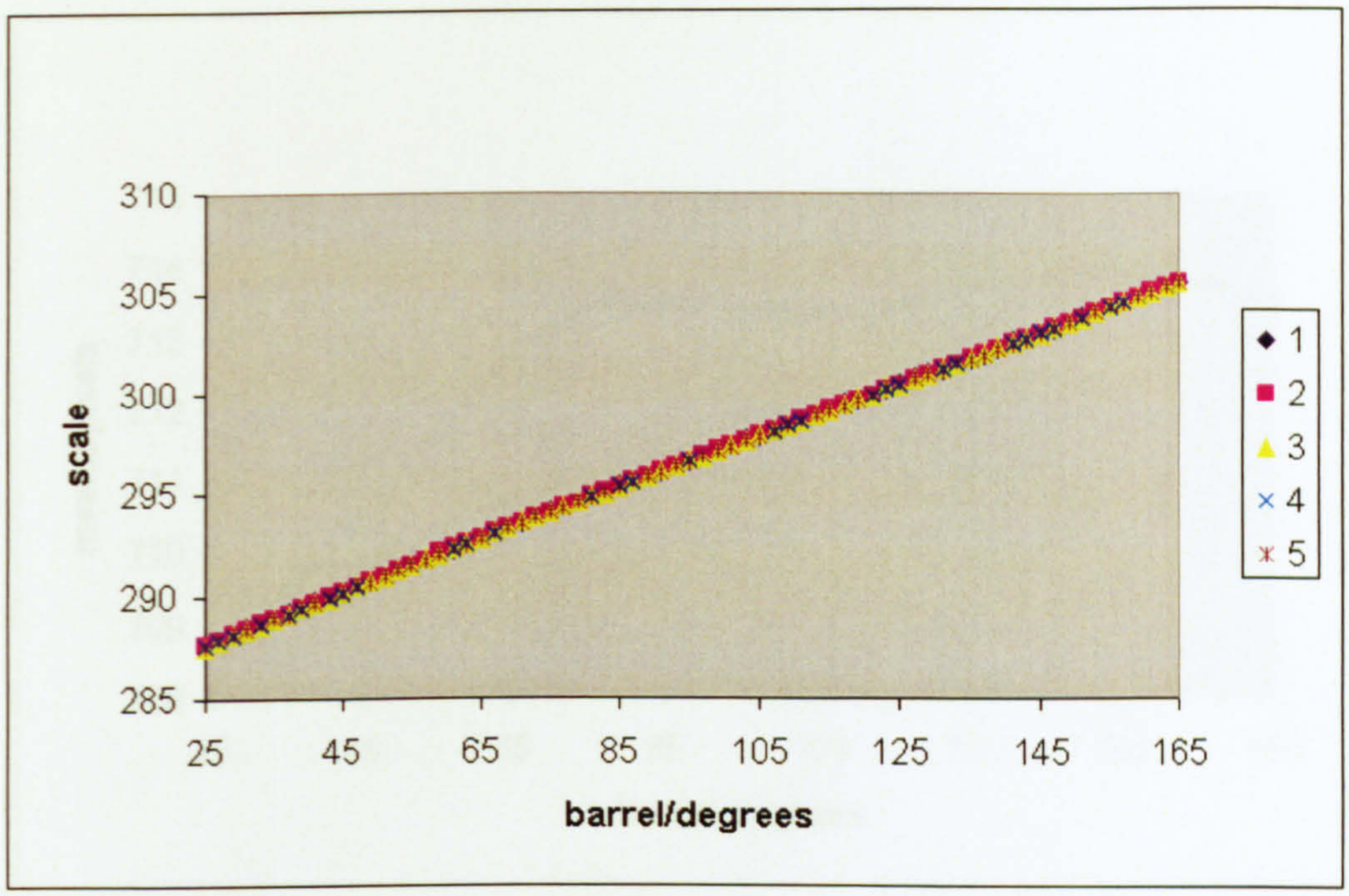

Figure 69 - the change in image scale with varying lens focus, left camera

The translations shown in Figure 67 and Figure 68 are generally repeatable between runs to within one pixel, and show no dominant simple behaviour. The scale change shown in Figure 69 shows high repeatability and a linear behaviour, as expected.

6.3.4 Characterisation – left lens

The positions of the target centre can be represented by an average of the five runs. The points in Figure 70 and Figure 71 below are the means of the five values for that focus position from the five data runs, and don't show any pattern not previously seen. Figure 72 plots the points on an x,y grid to illustrate the movement in two dimensions, and shows the complex nature of the image translations.

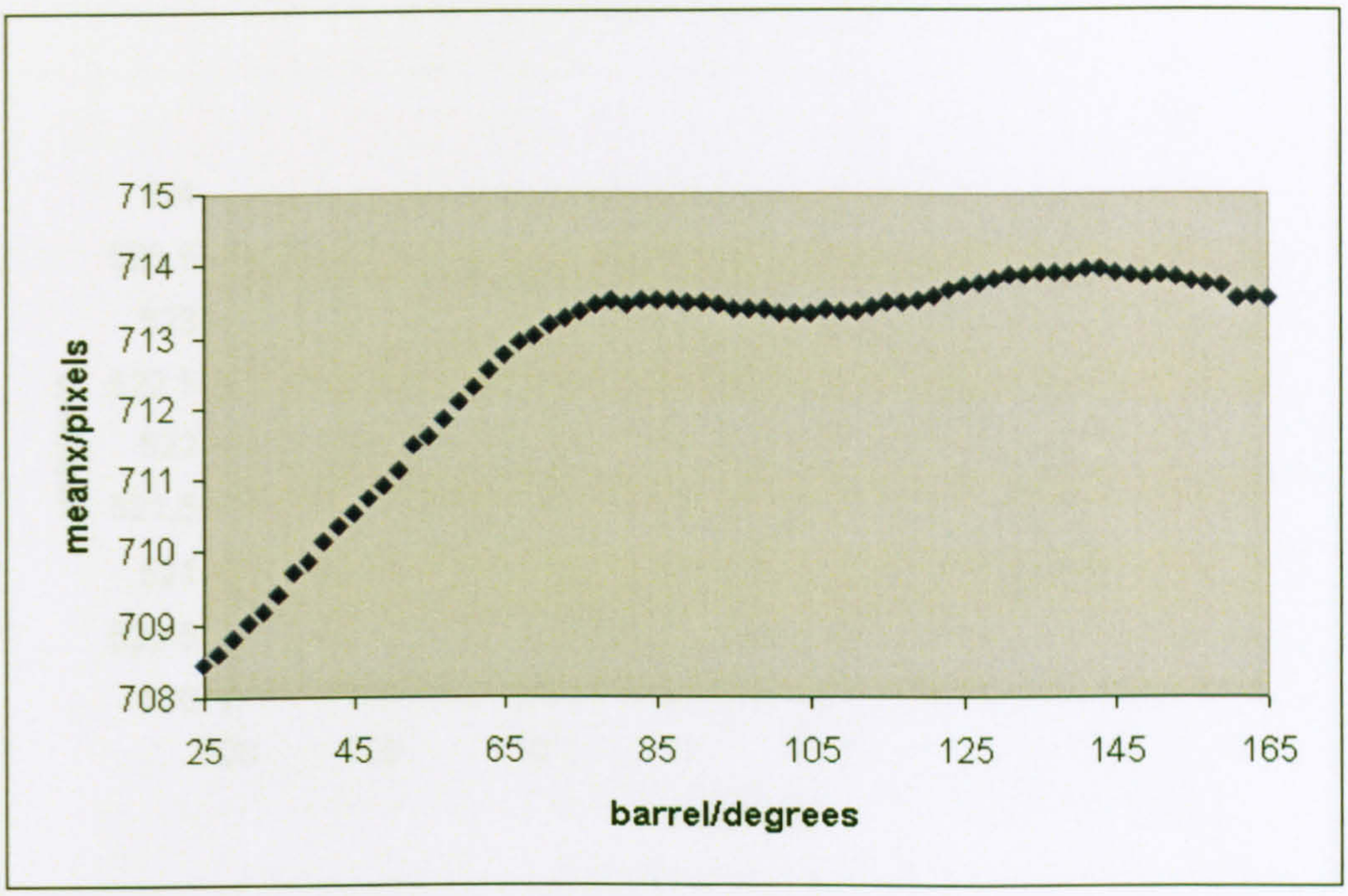

Figure 70 - mean target centre x-position xc with varying lens focus, left camera

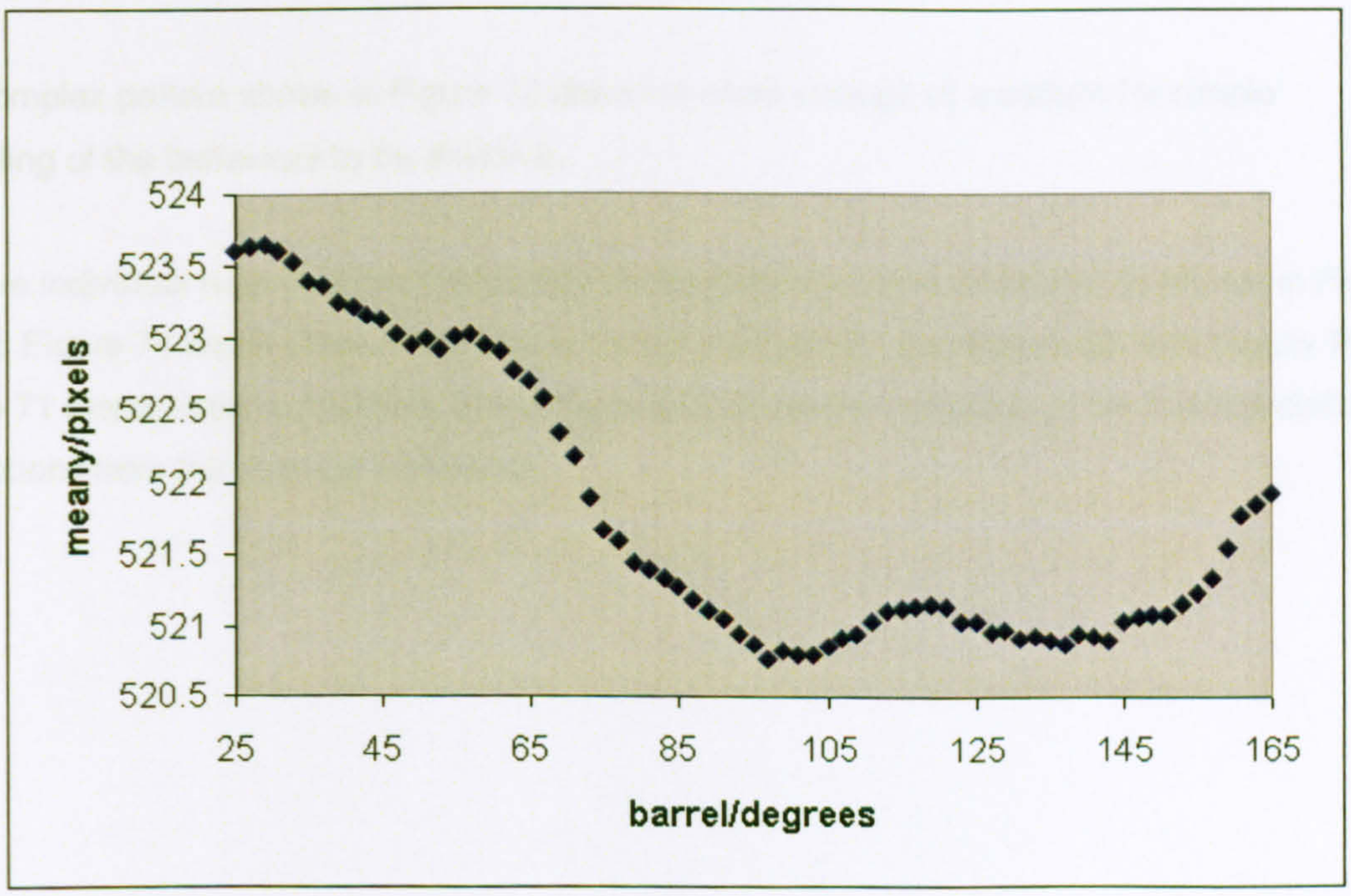

Figure 71 - mean target centre y-position yc with varying lens focus, left camera

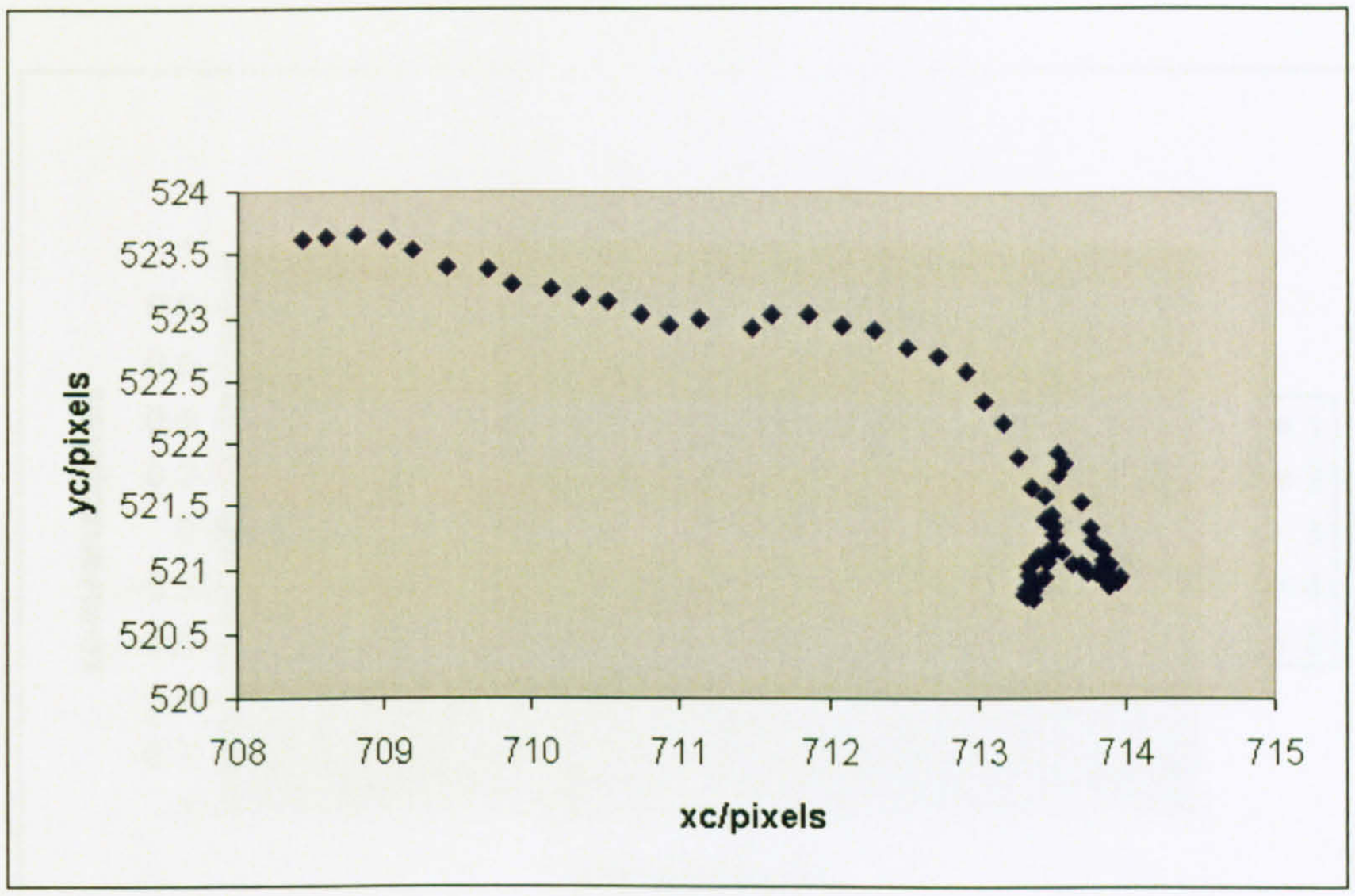

Figure 72 - movement of target centre (xc,yc) across the image as the lens is focused, left camera

The complex pattern shown in Figure 72 does not show enough of a pattern for simple modelling of the behaviour to be effective.

The five individual runs illustrate typical deviations from the mean positions, as shown in Figure 73 and Figure 74 below. The values are as those in Figure 67 and Figure 68, with Figure 70 and Figure 71 respectively subtracted. These figures illustrate the variability of the five translation calibrations from the average behaviour.

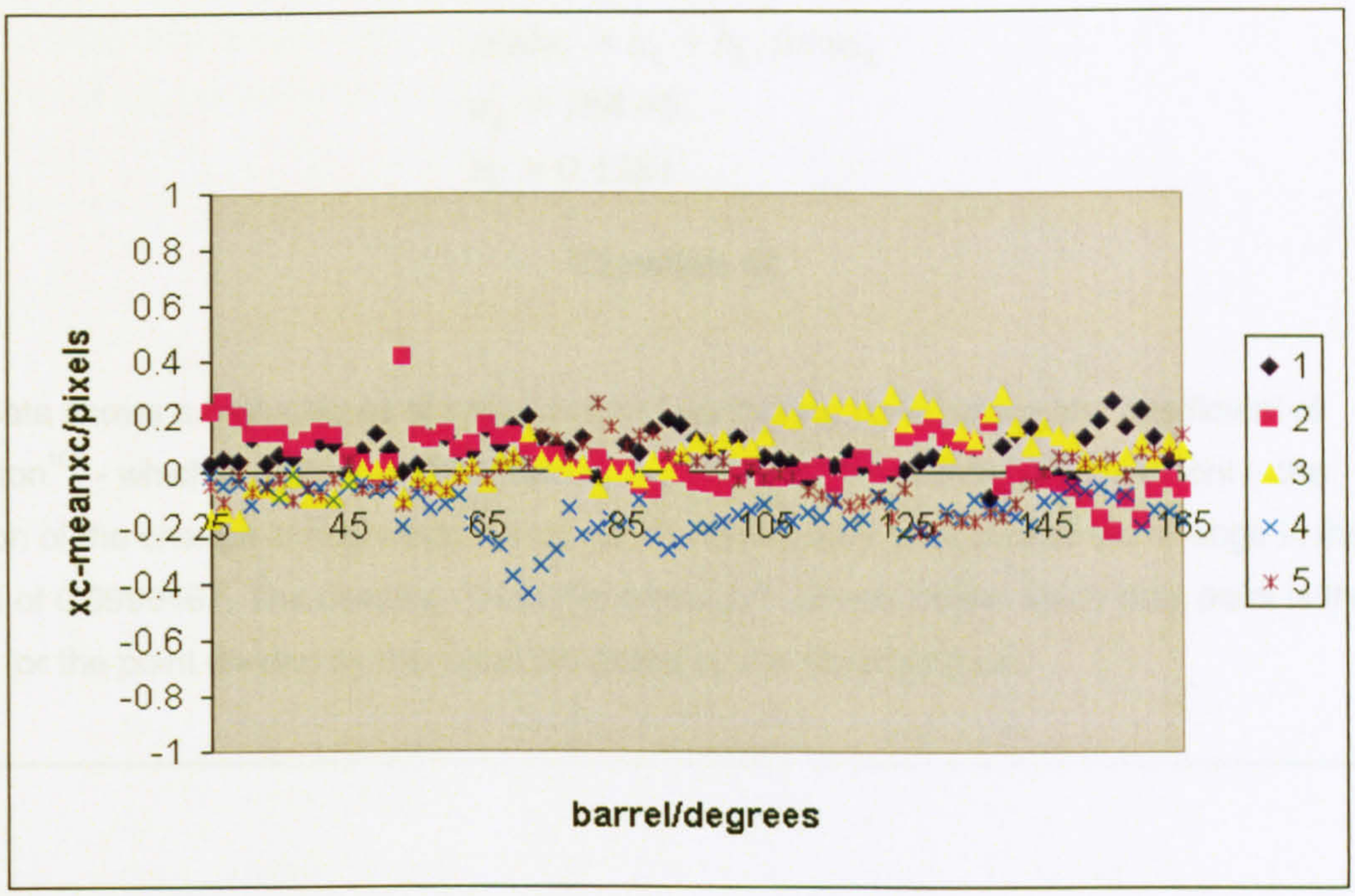

Figure 73 - deviation of target centre x-position xc from the mean with varying lens focus, left camera

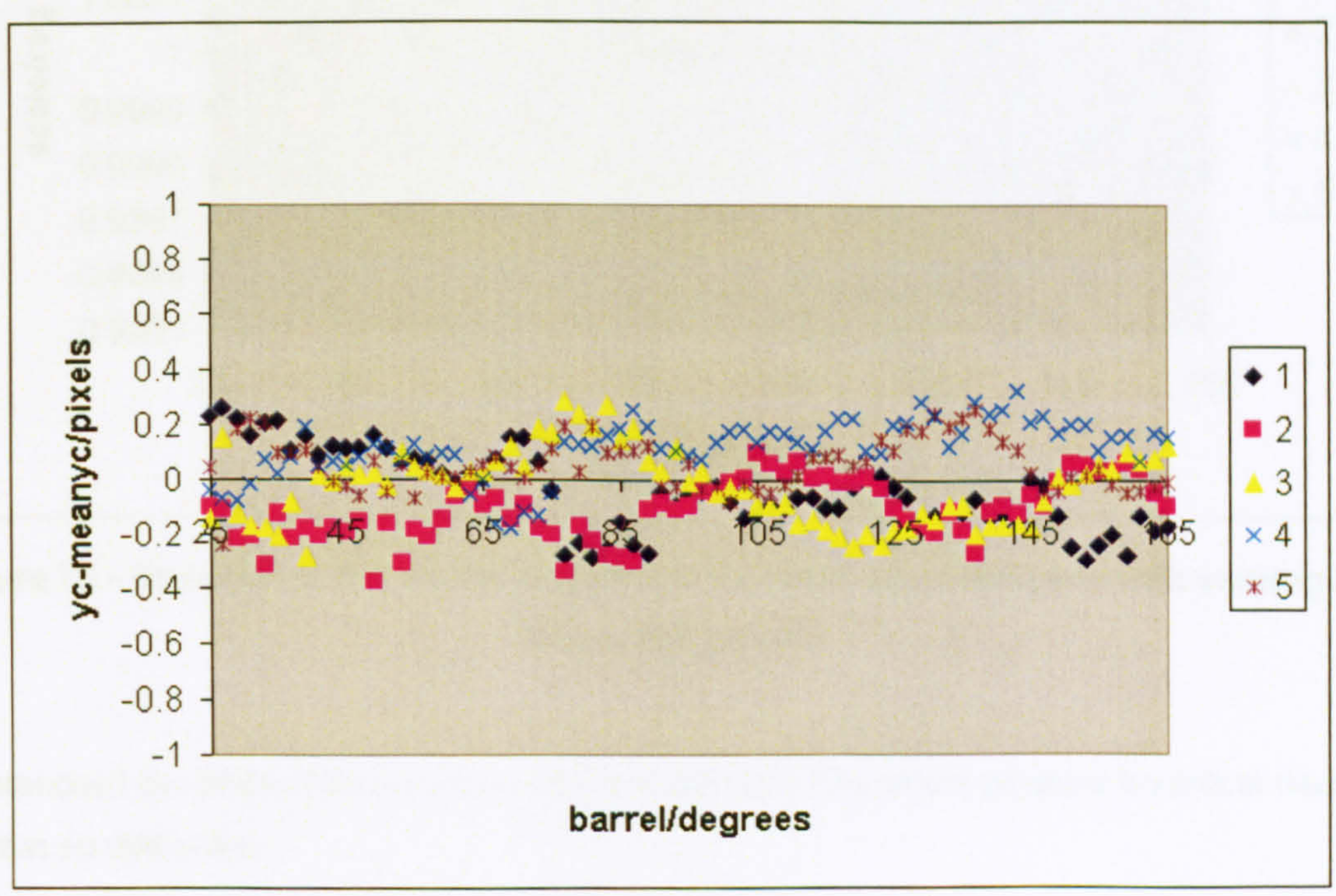

Figure 74 - deviation of target centre y-position yc from the mean with varying lens focus, left camera

The standard deviation from the mean is 0.124 in the x-direction, and 0.142 in the y-direction.

The zoom can be represented by a linear function obtained by least-squares fitting the image scale data. The function is as shown in Equation 44, and the constants shown are applicable only to the individual camera:

$$scale_L = a_L + b_L . focus_L$$
$$a_L = 284.46$$
$$b_L = 0.1281$$

Equation 44

The data shows a high degree of linear correlation, having an R^2 value (the coefficient of variation[106] - which is square of the Pearson product moment correlation coefficient – the fraction of the change in one variable that can be statistically attributed to the change in the other) of 0.9999467. The deviation from this function is shown below. Each data point is the scale for the point divided by the value predicted by the fitted function.

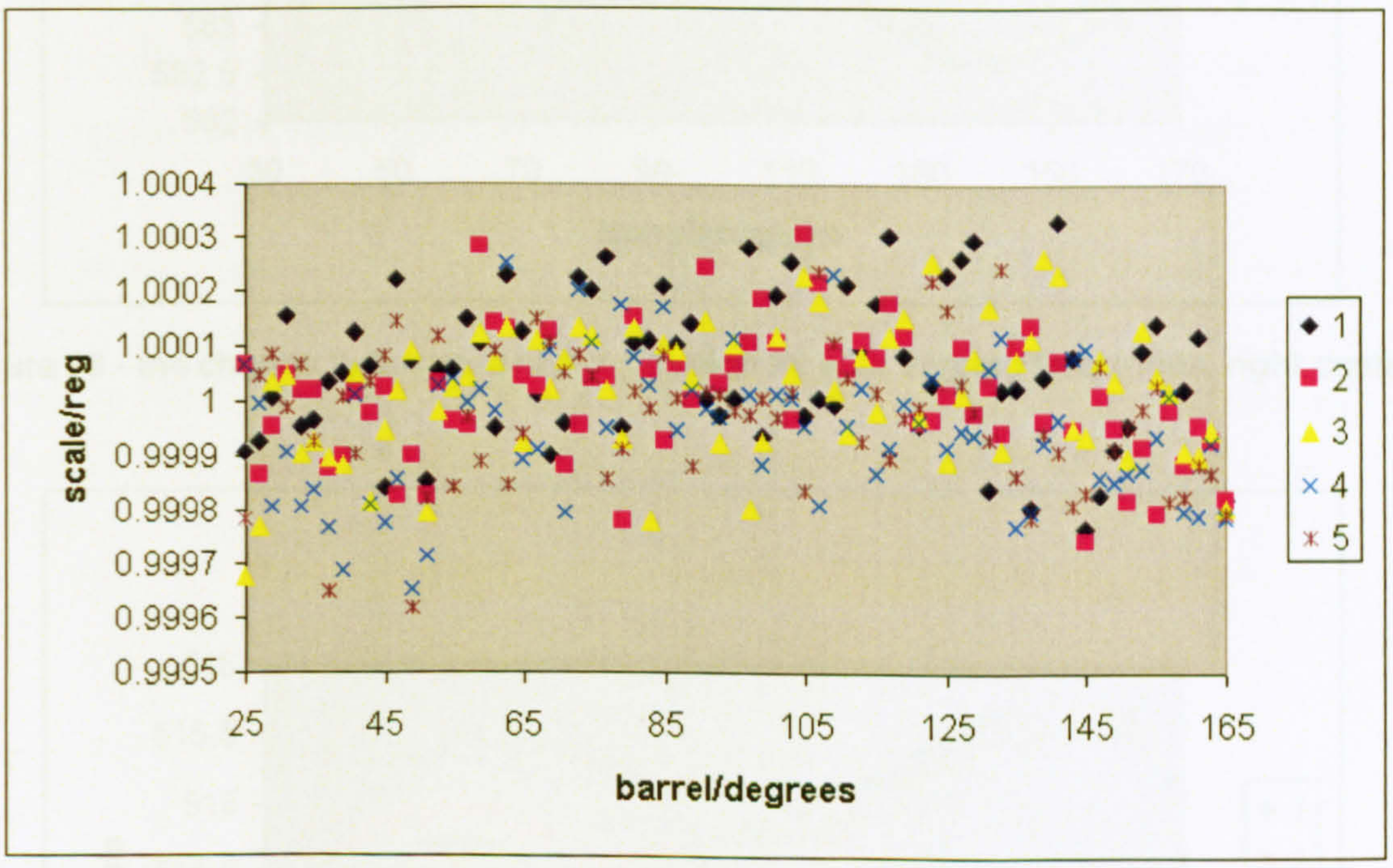

Figure 75 - deviation of the image scale from the fitted scale estimate with varying lens focus, left camera

The standard deviation of the scale spread is 0.000129. This would produce a vertical disparity of within ±0.066 pixels

6.3.5 Results – right lens

The right lens was calibrated, exactly as with the left lens, over a range of 140° (corresponding to focusing from beyond infinity to within 10cm), with a resolution of 2°, repeated four times to give five full sets of positions and scales. The results are shown in Figure 76, Figure 77 and

Figure 78. Figure 76 and Figure 77 show the translation of the target image points' centre of mass as a function of lens focus, and Figure 78 shows the changing scale of the target image, also as a function of the focus position of the lens.

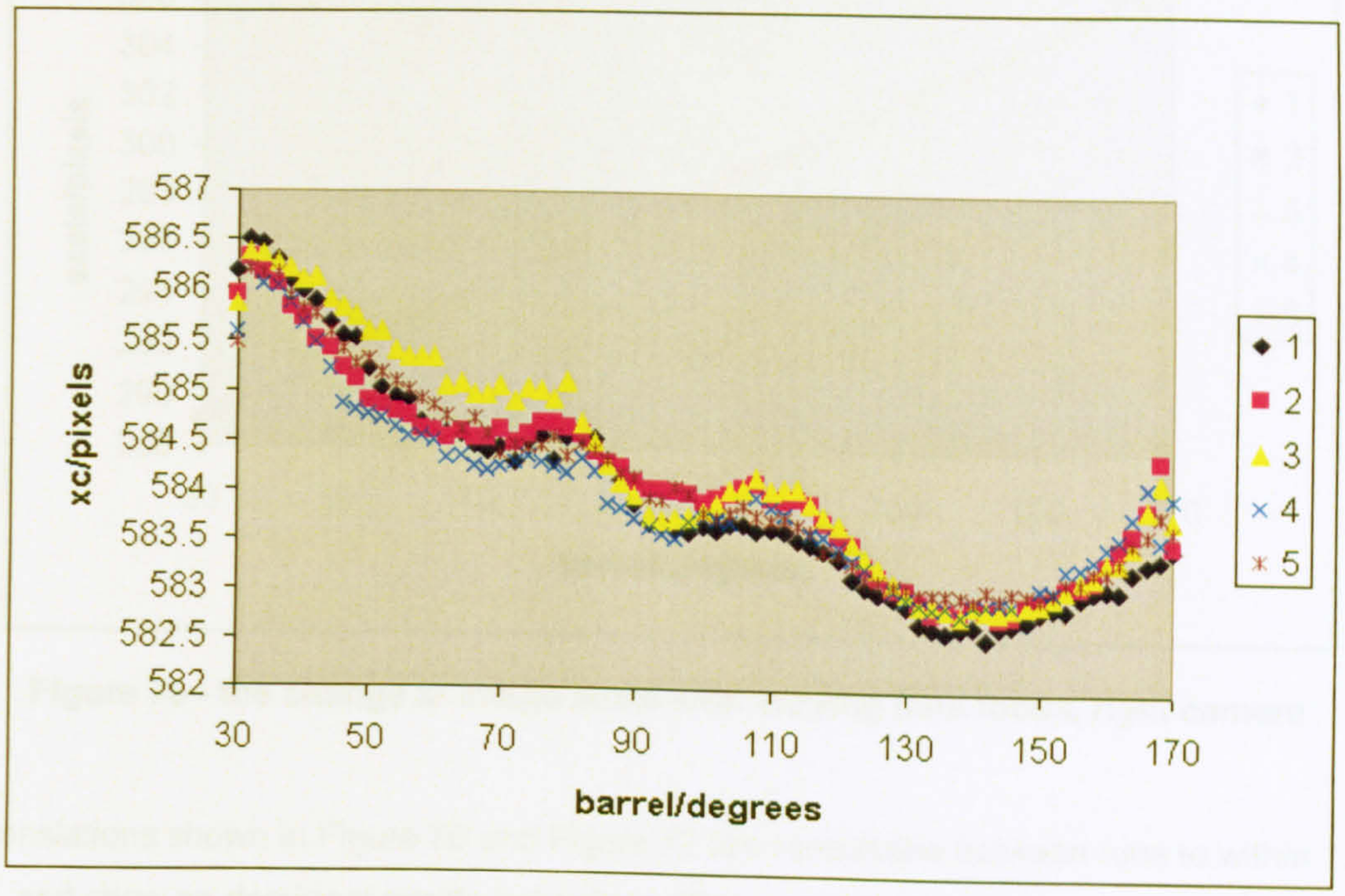

Figure 76 - the change in target centre x-position xc with varying lens focus, right camera

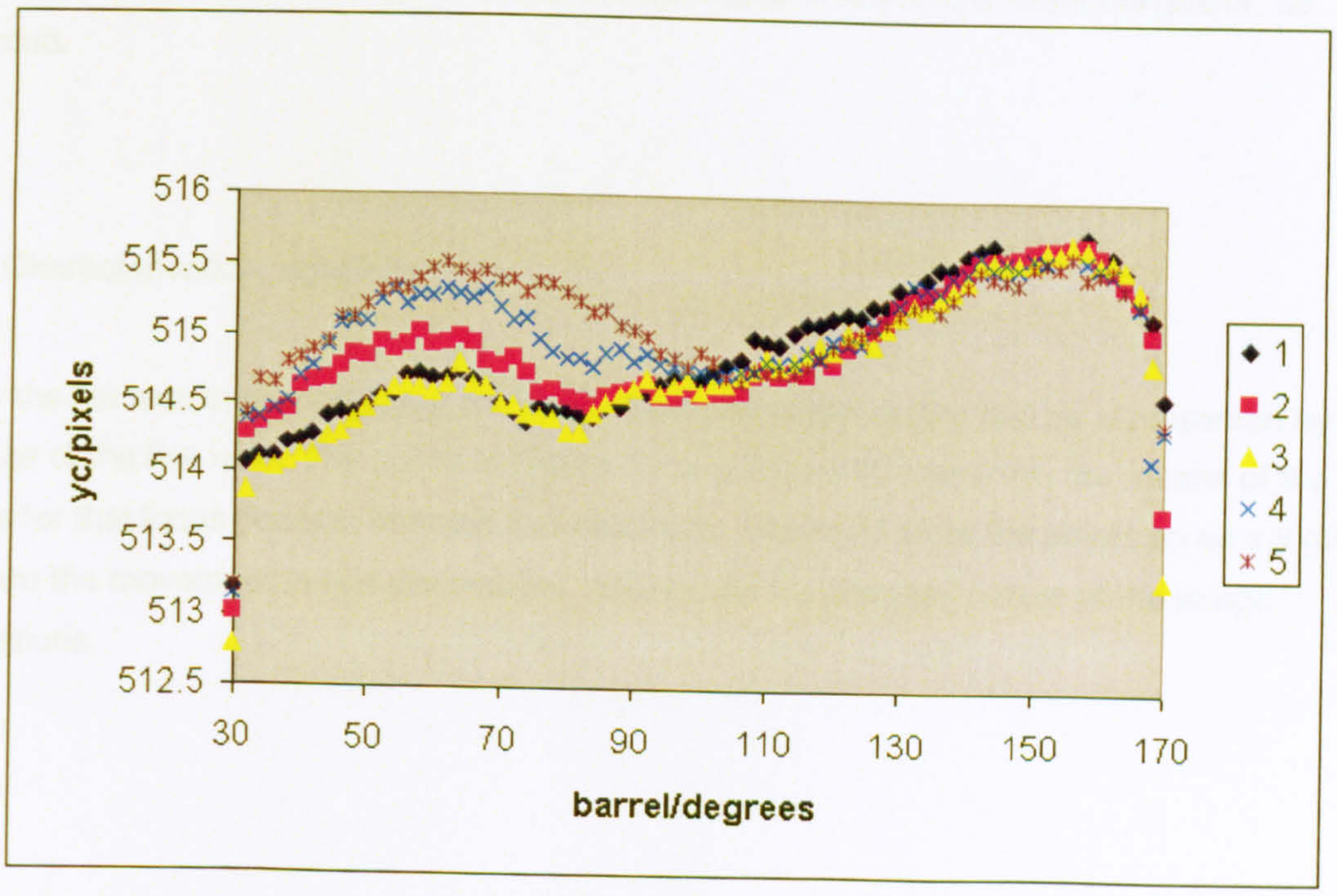

Figure 77 - the change in target centre y-position yc with varying lens focus, right camera

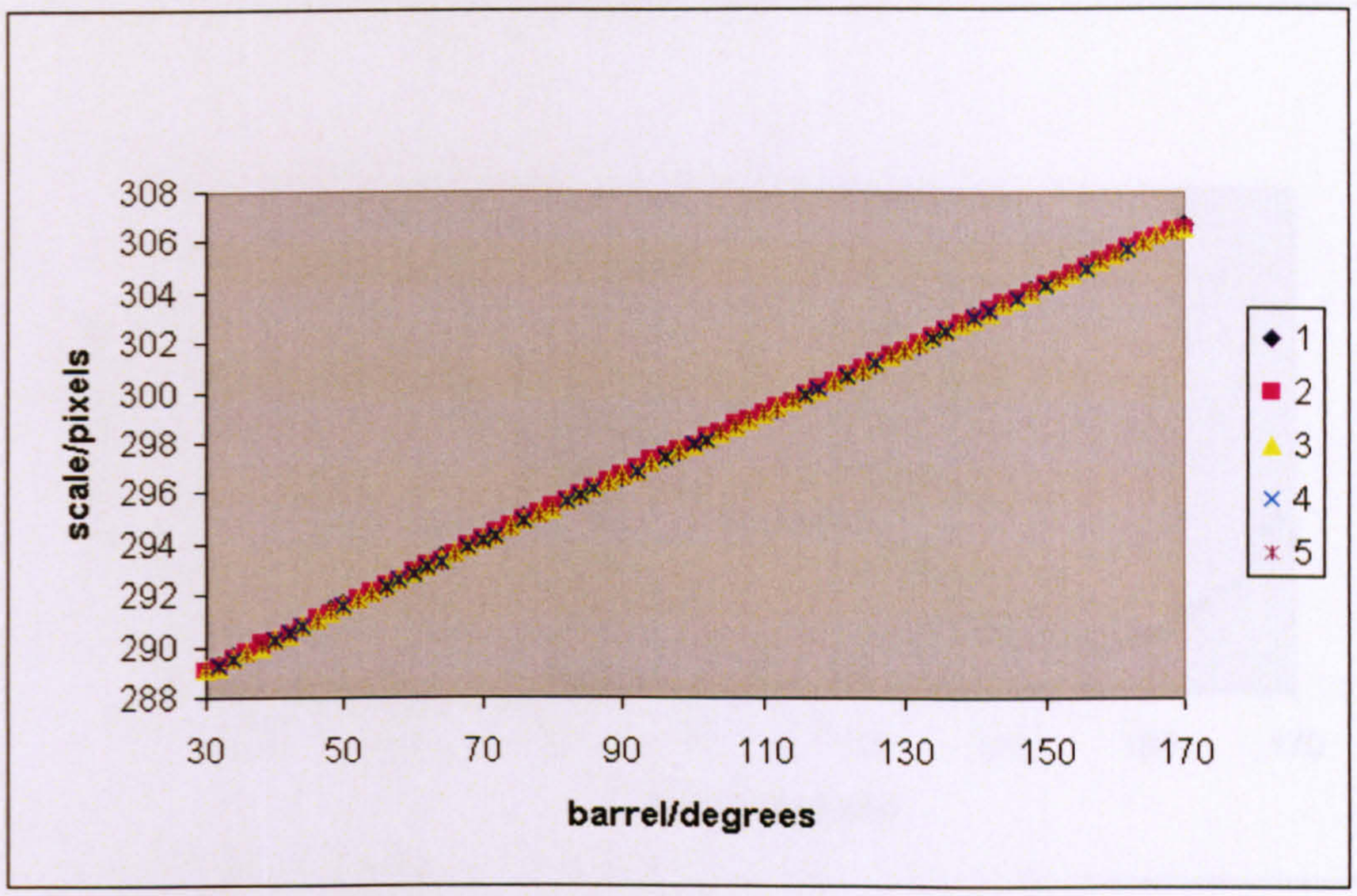

Figure 78 - the change in image scale with varying lens focus, right camera

The translations shown in Figure 76 and Figure 77 are repeatable between runs to within 1.1 pixels, and show no dominant simple behaviour. The y-position data varies much more from 30° to 100° than from 100° to 170°, possibly due to a loose component within the lens assembly. The scale change shown in Figure 78 shows high repeatability and a linear behaviour, as expected.

6.3.6 Characterisation, right lens

As for the left lens in section 6.3.4, the positions of the target centre can be represented by an average of the five runs. The points in Figure 79 and Figure 80 below are the means of the five values for that focus position from the five data runs. Figure 81 plots the points on an x,y grid to illustrate the movement in two dimensions, and shows the complex nature of the image translations.

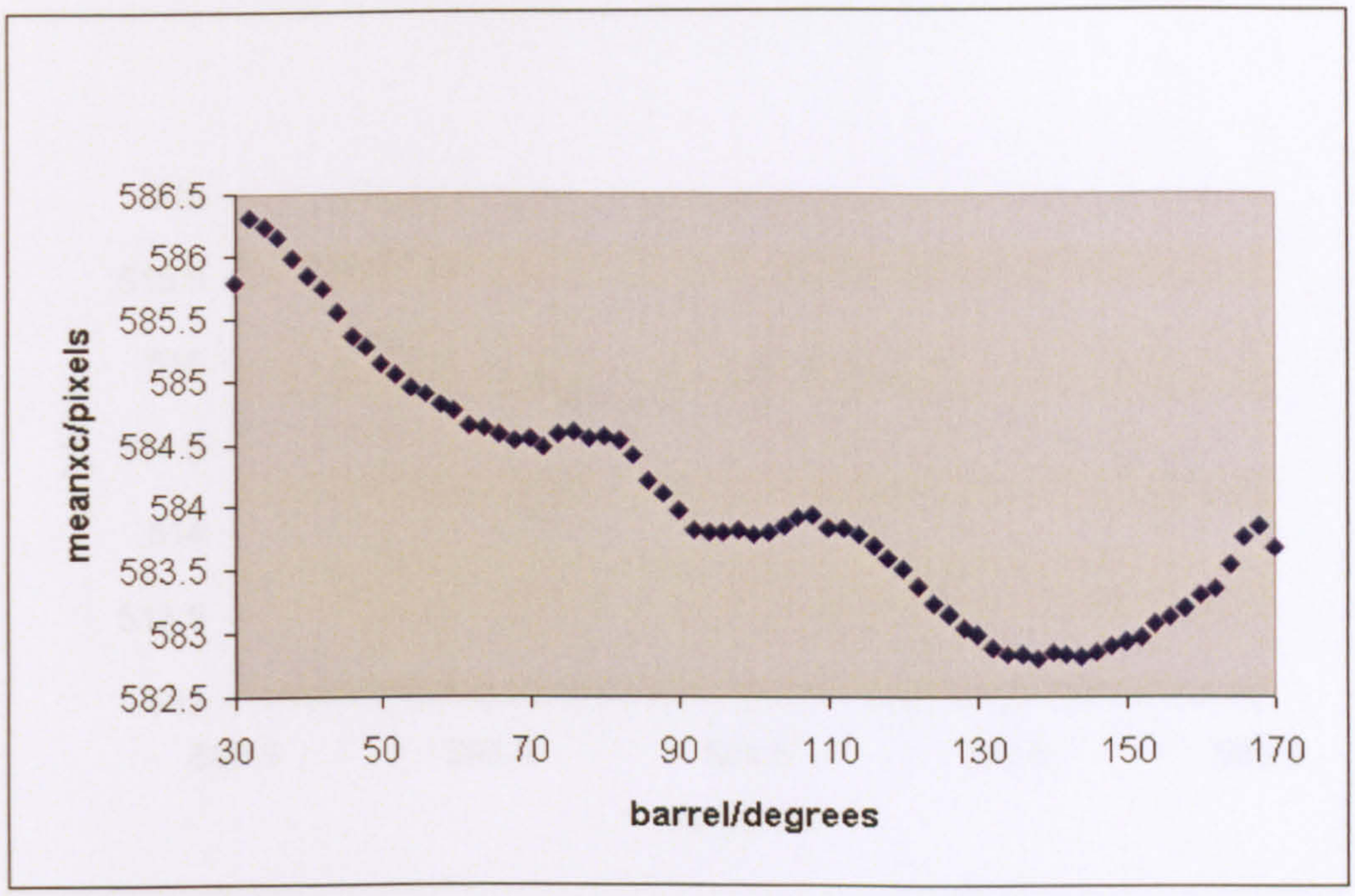

Figure 79 - mean target centre x-position with varying lens focus, right camera

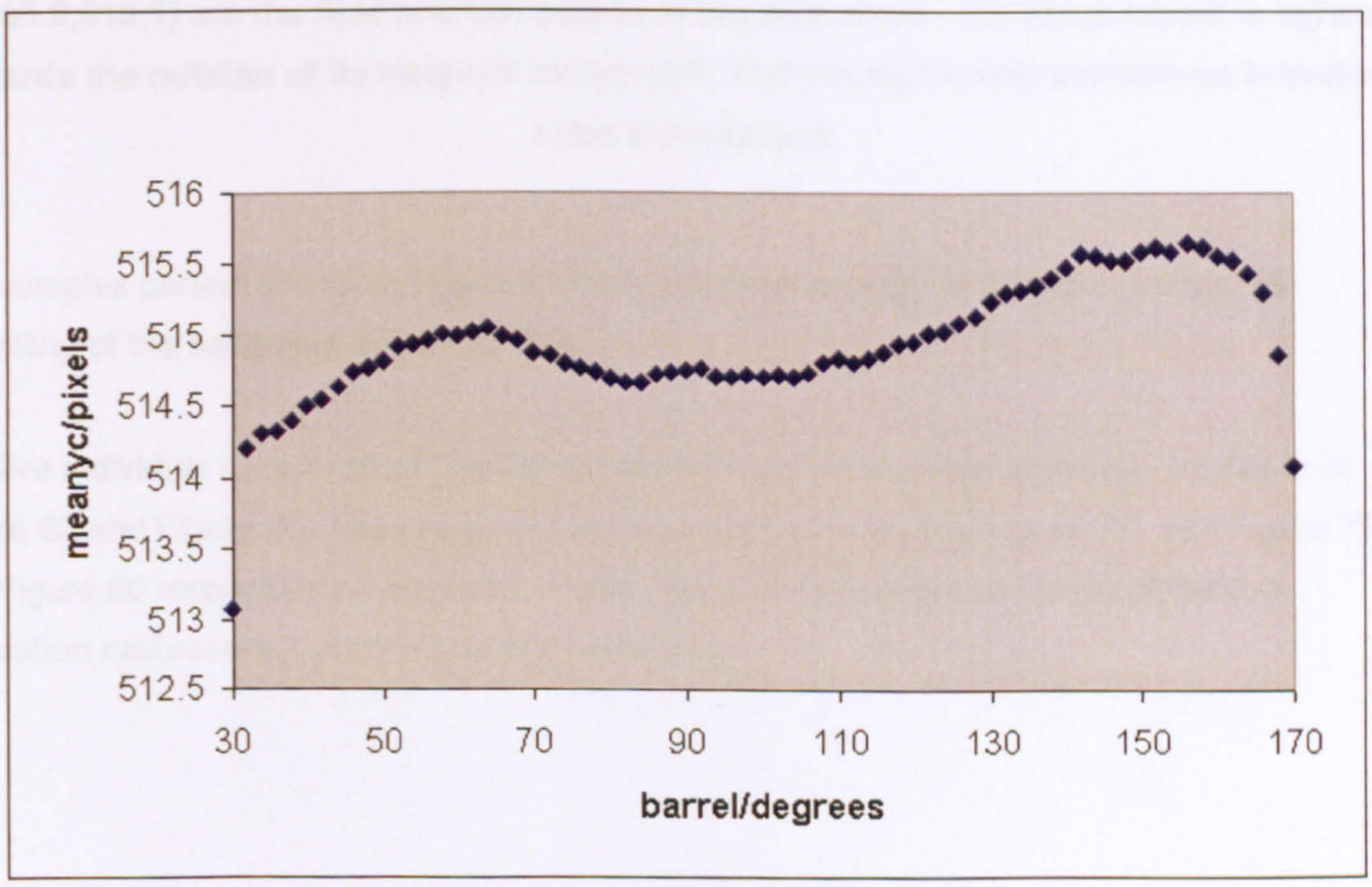

Figure 80 - mean target centre y-position with varying lens focus, right camera

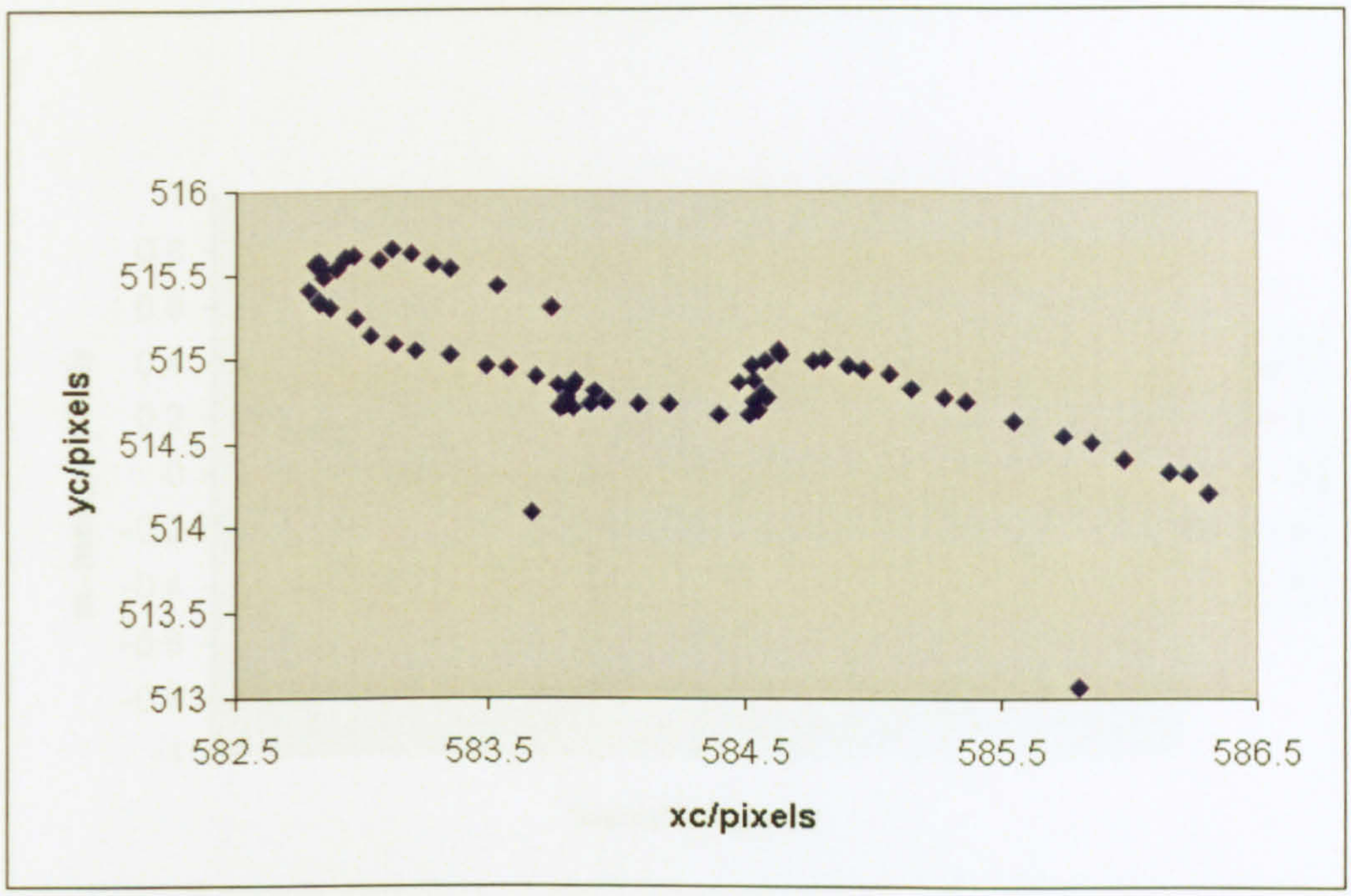

Figure 81 - movement of target centre across the image as the lens is focused, right camera. The two points located away from the rest of the pattern (at 583.7,514.1 and 585.8,513.1) are the first and last points in the sequence - the focus barrel is tighter towards the outside of its range of movement, and the tightening sometimes introduces extra translations

The complex pattern shown in Figure 81 does not show enough of a pattern for simple modelling of the behaviour to be effective.

The five individual runs illustrate typical deviations from the average positions, as shown in Figure 82 and Figure 83. The values are as those in Figure 76 and Figure 77, with Figure 79 and Figure 80 respectively subtracted. These figures illustrate the variability of the five translation calibrations from the average behaviour.

Figure 82 – deviation of target centre x-position from the mean with varying lens focus, right camera

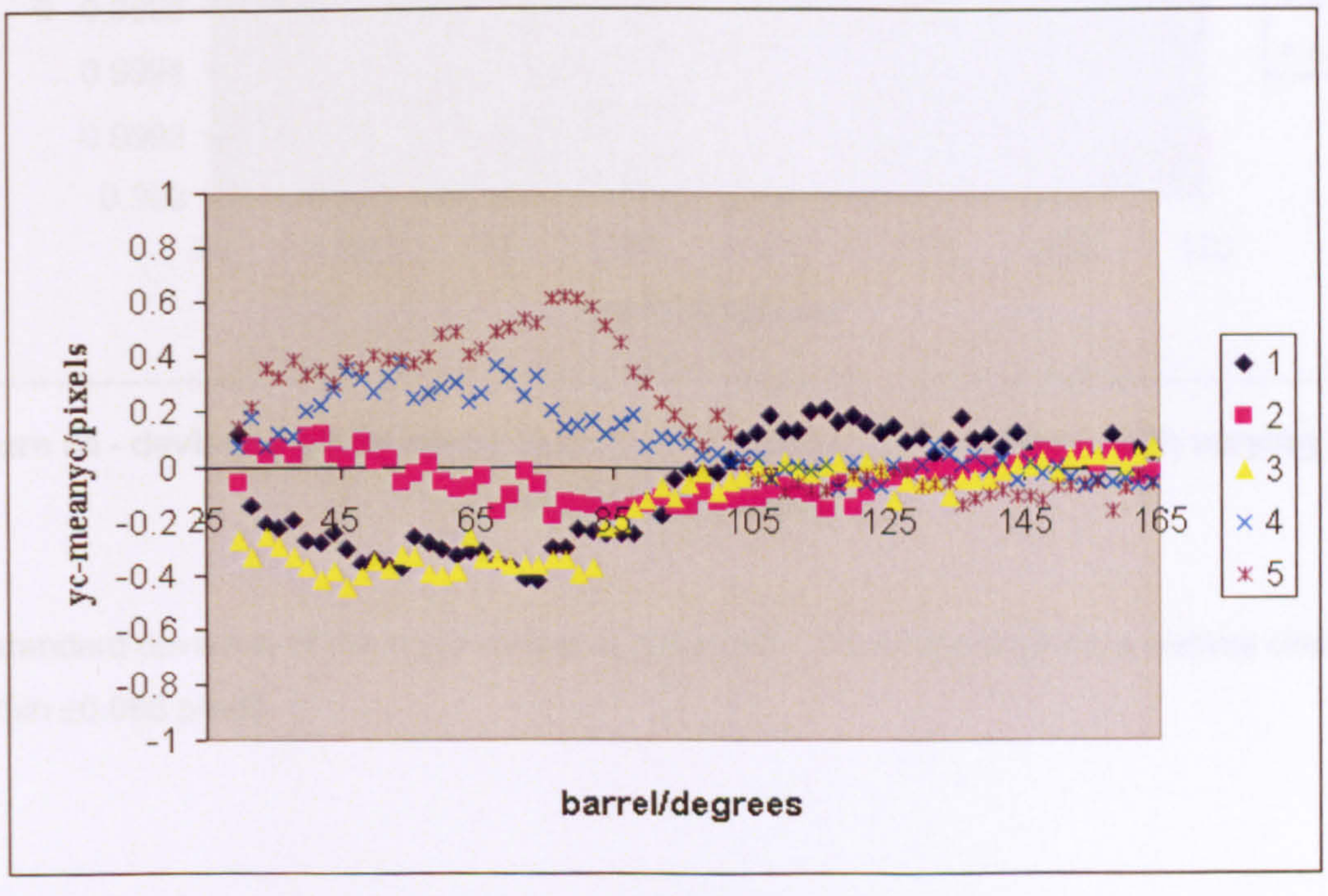

Figure 83 - deviation of target centre y-position from the mean with varying lens focus, right camera

The standard deviation from the mean is 0.190 in the x-direction, and 0.220 in the y-direction.

As with the left camera, the zoom is represented by a linear function obtained by least-squares fitting the image scale data. The function is given in Equation 45, and the values for the coefficients are specific to the camera measured.

$$scale_R = a_R + b_R.focus_R$$
$$a_R = 285.24$$
$$b_R = 0.1267$$

Equation 45

The data shows a high degree of linear correlation, having an R^2 value of 0.9999068. The deviation from this function is shown in Figure 84. Each data point is the scale for the point divided by the value predicted by the fitted function.

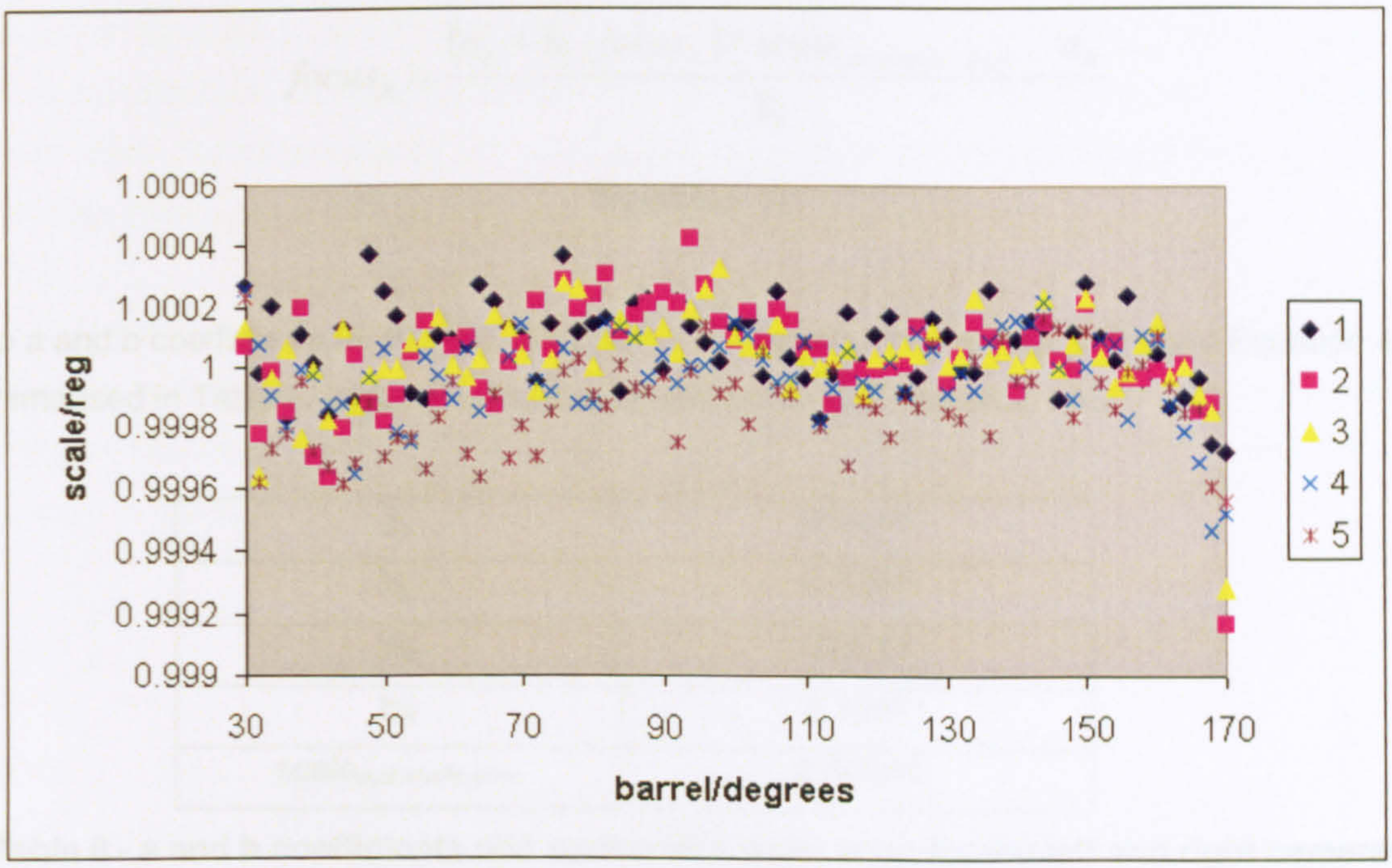

Figure 84 - deviation of the image scale from the fitted scale estimate with varying lens focus, right camera

The standard deviation of the scale spread is 0.000168. This would produce a vertical disparity of within ±0.086 pixels

6.3.7 Scale matching

By combining the information about each of the two cameras, the behaviour of the camera pair can be determined. The first priority is to match the scales of the images produced by the cameras. For this, Equation 44 and Equation 45 must equate as shown in Equation 46.

$$scale_R = scale_L * scale_{systmatic-error}$$
$$a_R + b_R.focus_R = (a_L + b_L.focus_L) * scale_{systematic-error}$$

Equation 46

The scales must be equal, taking into account the systematic error described in section 6.1.10.

Rearranging gives:

$$focus_R = \frac{(a_L + b_L.focus_L) * scale_{systematic-error} \quad a_R}{b_R}$$

Equation 47

The a and b coefficients for the two calibrated cameras are, from Equation 44 and Equation 45, summarised in Table 8, along with the systematic error from section 6.1.10:

a_L	284.46
b_L	0.1281
a_R	285.24
b_R	0.1267
$scale_{systematic-error}$	0.99984

Table 8 - a and b coefficients and systematic scale error for the left and right cameras

Substituting values from Table 8 into Equation 47 gives:

$$focus_R = 1.0109 * focus_L \quad 6.52$$

Equation 48

When using the camera, the focus can now be set using the left lens, and Equation 48 can be used to determine the correct focus to set the right lens to, in order to match the scales of the images.

6.3.8 Translation estimates

Using the motion of the target centre in the images taken by the two cameras, the translation between the images can be estimated. The translation from the left image to the right image is

simply the target centre position in the right image minus the target centre position in the left image, at focus positions producing equal scales in the two images.

Figure 85 and Figure 86 show the translation between the two cameras for the previously measured values of focus position. To calculate the target centre position for the right lens, it is necessary to interpolate between the measured values, because the right lens focus positions are no longer restricted to the previously measured values. For example, if the left lens is focused at 123°, Equation 48 gives the required right lens focus position as 118.4°. The calibration measured the target centre position at 118° and 120°, so a simple linear interpolation means that the target centre position at 118.4° can be calculated as 0.8 of the position at 118° plus 0.2 of the position at 120°. In practice, this interpolation needs to be performed for both lenses since the left lens position does not have to be set to one of the previously measured values either, but for purposes of illustration here it is simpler to show translations using the previously measured values for only one lens.

The range of operation of the cameras as a stereo pair is slightly smaller than for each lens calibration individually, since the range of the pair may only include those focus positions where the scales can be matched between the two cameras. This reduced range still includes the entire set of focus positions needed for the stereo camera's use.

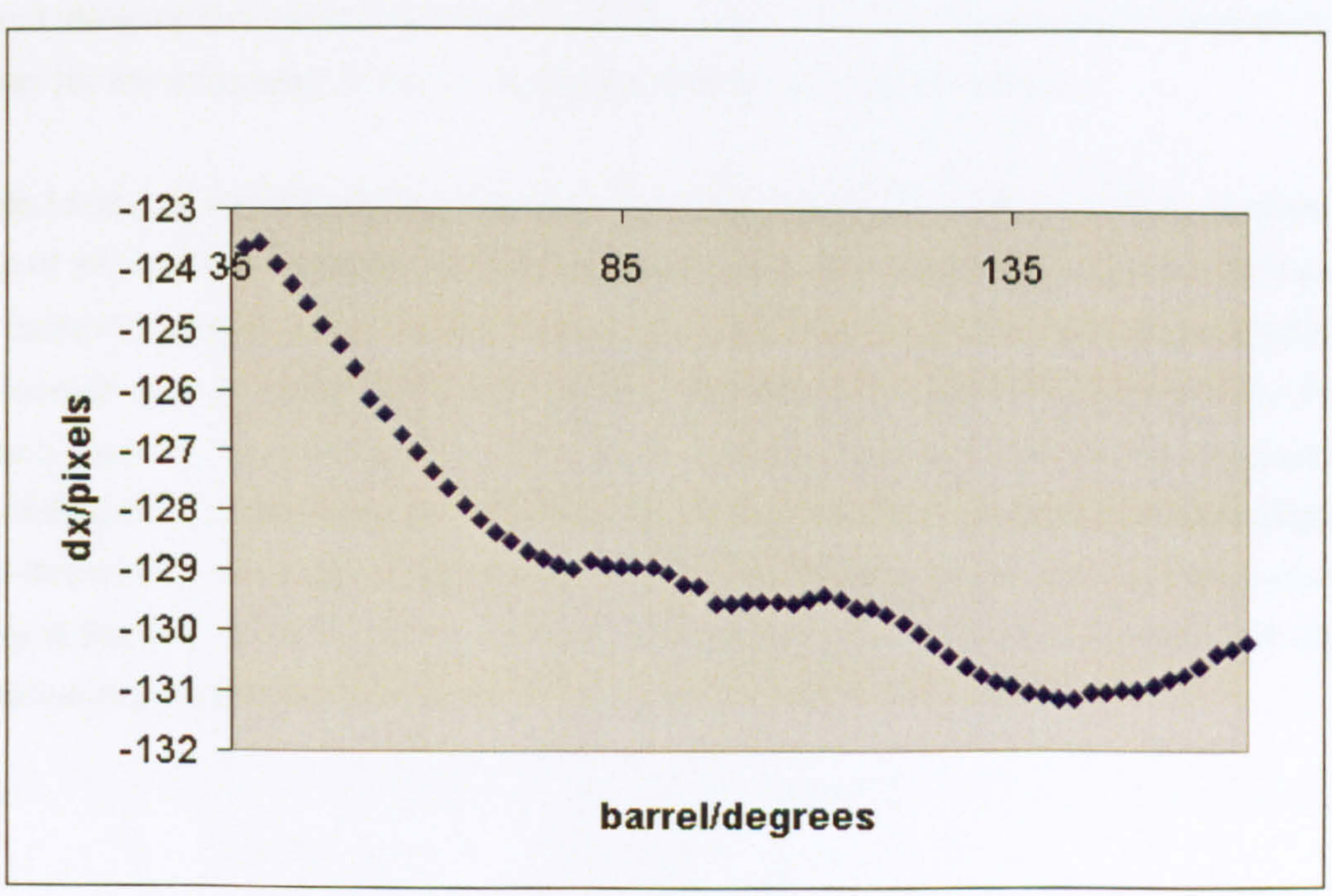

Figure 85 – x-translation dx from the left to the right image with varying left lens focus

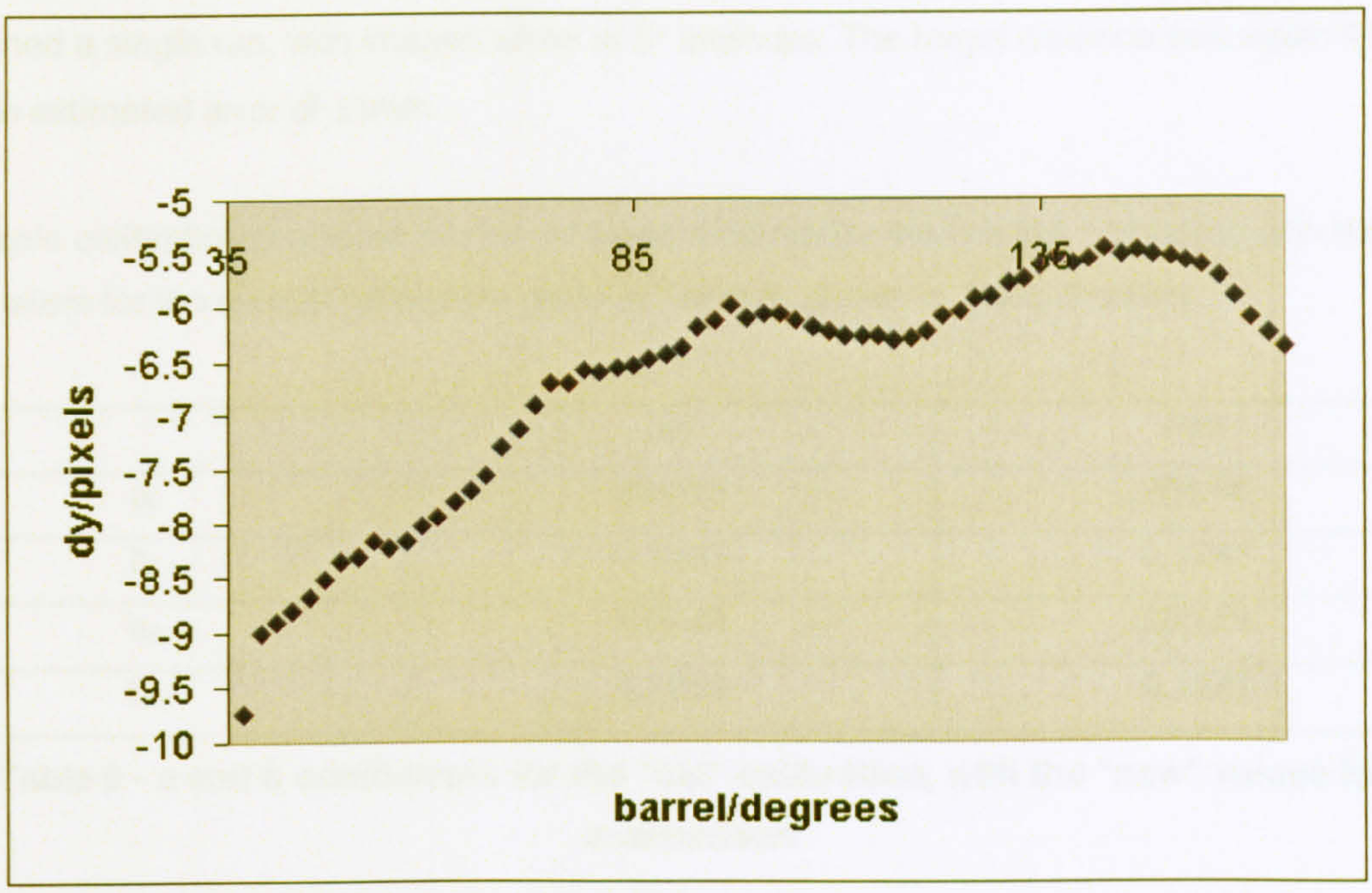

Figure 86 - y-translation dy from the left to the right image with varying left lens focus

The above two figures show an estimate for the translations between images taken with the stereo camera. The x-translation is largely dependent on the distance to the calibration target (90cm) and the horizontal separation of the cameras (10cm), and most of that translation is due to normal stereoscopic horizontal disparity. In this case, and with camera parameters as in Equation 10, the x-translation due to horizontal disparity is -132.67 pixels.

The data for dx, dy can be used to crop an image pair taken with the stereoscopic camera, as a function of the lens focus positions, after including a correction for the systematic error caused by the calibration target, as described in section 6.1.10. This correction is simply an addition of the translation errors to the predicted dx and dy. The values for vertical disparity are the most important, because vertical disparity should be reduced as much as possible, so the cropping should always be the same as dy. The dx values may be used with an offset representing the desired distance at which the scene should exhibit zero disparity (since they represent zero disparity at 90cm from the camera - the horizontal disparity due to horizontal separation is given by Equation 26), to set the ZDP distance as a function of lens focus positions.

6.3.9 Durability of the calibration

As noted in section 3.4, lens systems are generally unstable. A calibration of the stereo camera's lenses was made approximately six weeks before the detailed calibration above, so that the effect of time and use on the camera could be measured. During the interval, the camera assembly was gently moved to a few different locations in the office, and images were taken, involving adjustments being made to the lens focus positions. The initial calibration

contained a single run, with images taken at 5° intervals. The target distance was again 91cm, with an estimated error of ±1mm.

The scale calibration produced the following parameters for the line fits, compared with the parameters for the current calibration given in Table 8, shown in Table 9 below:

	old	new
a_L	284.65	284.46
b_L	0.1280	0.1281
a_R	285.48	285.24
b_R	0.1259	0.1267

Table 9 - a and b coefficients for the "old" calibration, with the "new" values for comparison

The R^2 values for the left and right line fits were 0.9999558 and 0.9998988 respectively. Deviations of the data from the fitted function are shown in Figure 87 and Figure 88.

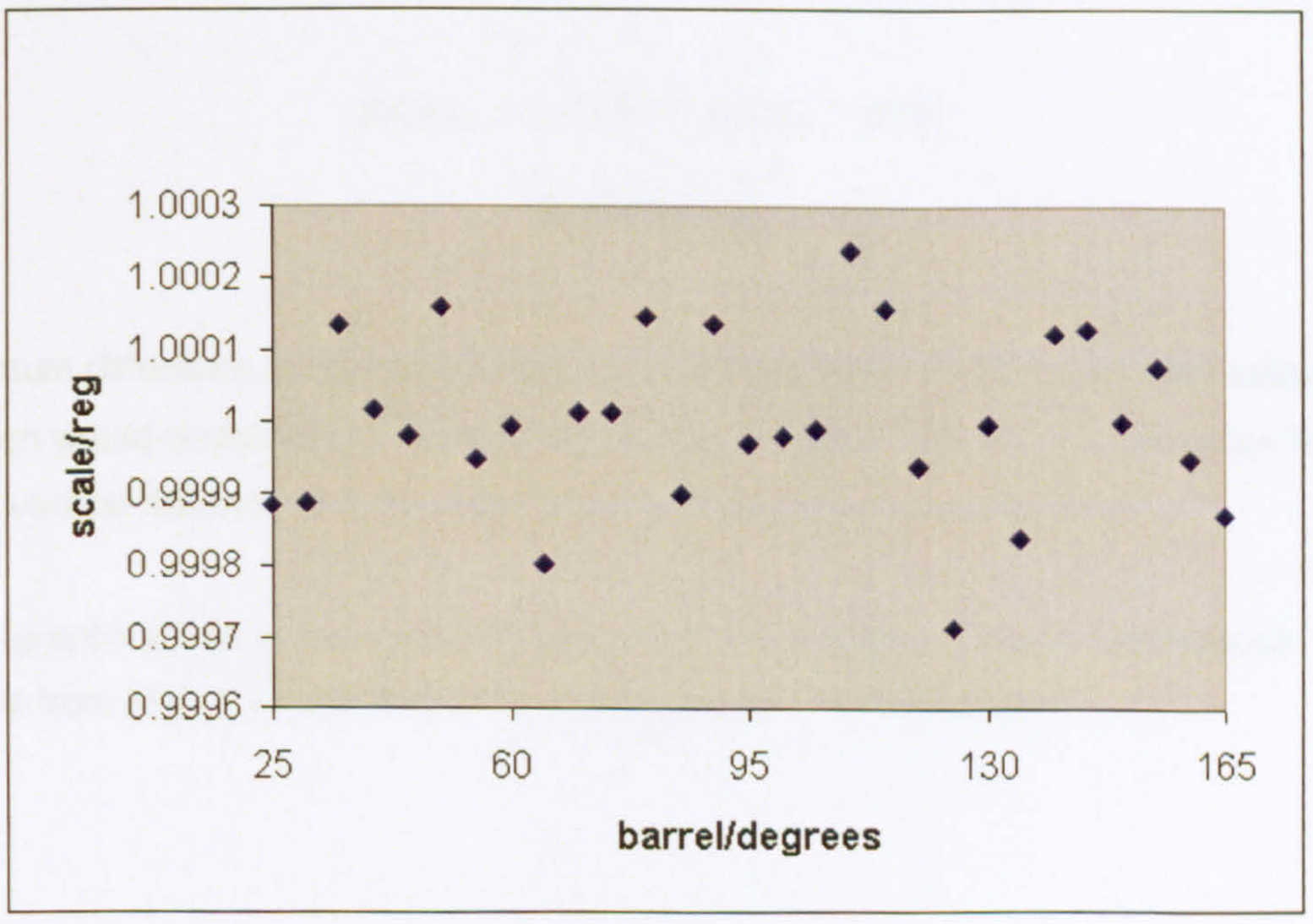

Figure 87 – deviation of the image scale from the fitted scale estimate with varying lens focus, left camera, old calibration

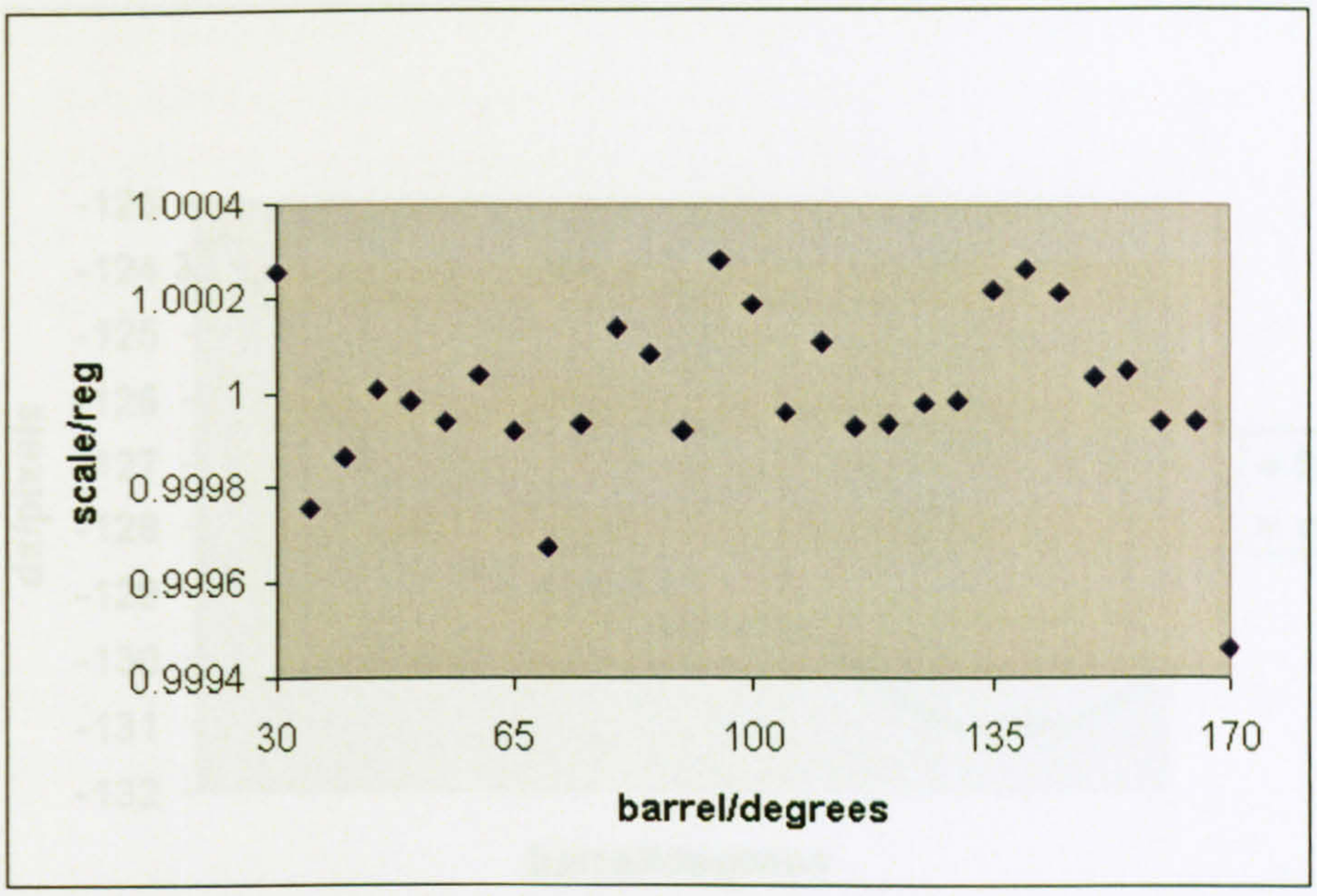

Figure 88 - deviation of the image scale from the fitted scale estimate with varying lens focus, right camera, old calibration

Equating the scales of the left and right images for the old calibration as in Equation 47 gives:

$$focus_R = 1.0169 * focus_L - 6.66$$

Equation 49

The maximum difference in right lens focus positions predicted by the full and old calibrations is 0.58°, which would correspond to a scale difference of about 0.024% (which equates to a maximum vertical disparity of 0.12 pixels) if the old calibration was in error.

Below are graphs showing the translations from the left to the right image, with the full calibrations from Figure 85 and Figure 86 superimposed for comparison.

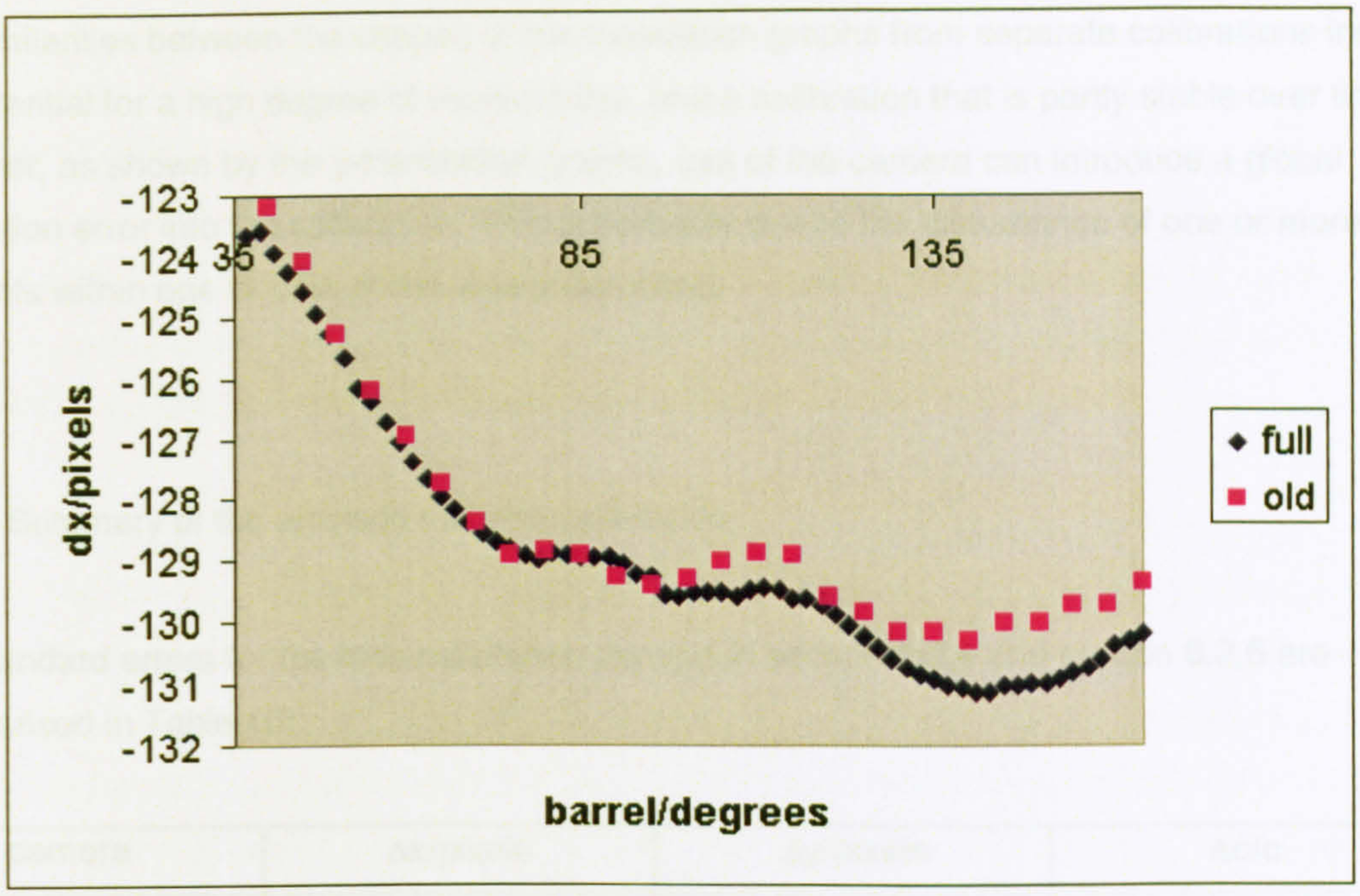

Figure 89 - x-translation from the left to the right image with varying left lens focus, both full and old calibrations

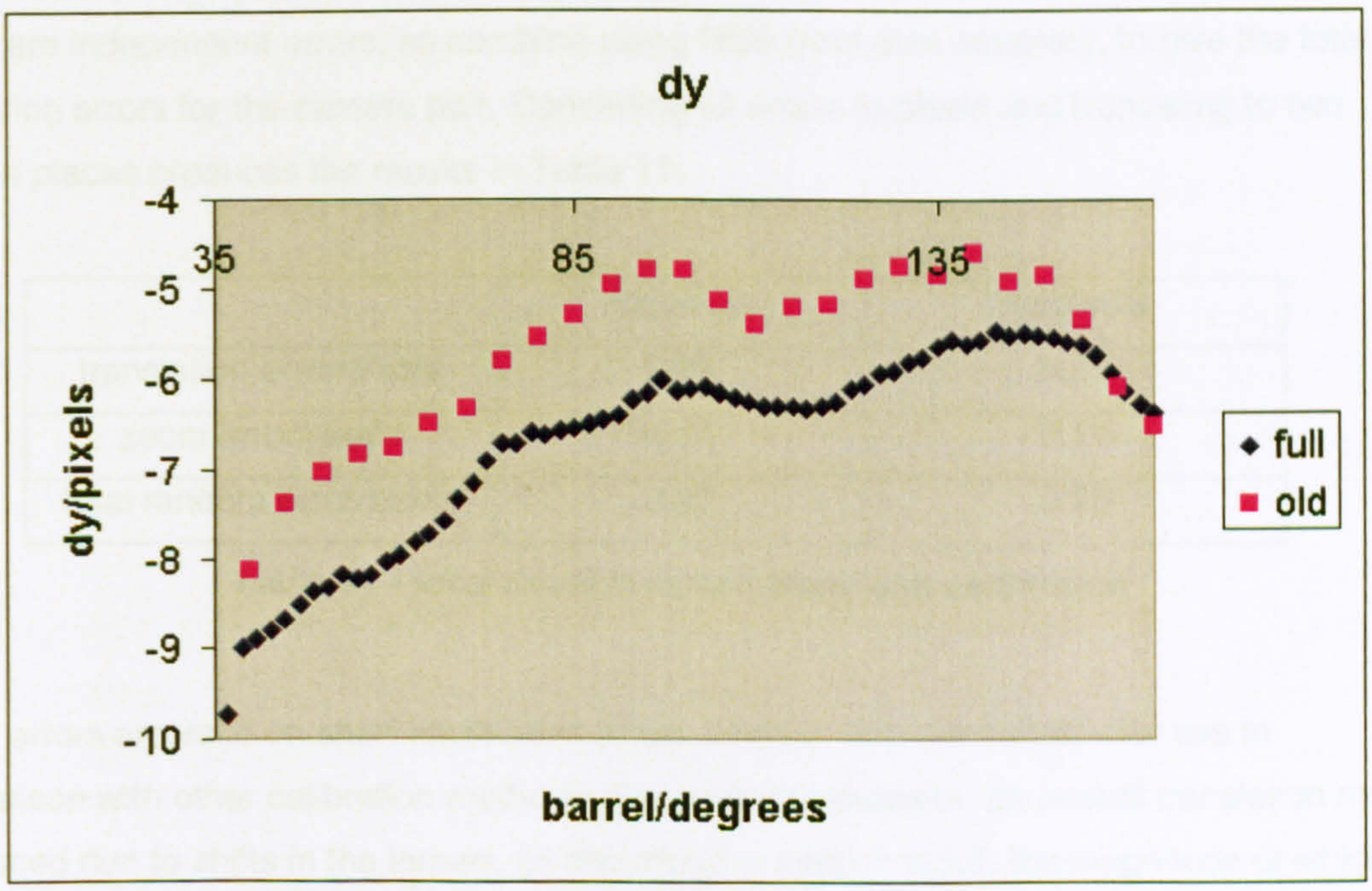

Figure 90 - y-translation from the left to the right image with varying left lens focus, both full and old calibrations

The x-translations are consistent to within 1.1 pixels, with most of the deviation being at the higher values of left lens focus position. The basic shape of the graphs is generally similar between the two calibrations. The y-translations are consistent to within 1.5 pixels, with a gap of 1 to 1.5 pixels being maintained at the majority of focus positions. Again, the basic shape of the graphs are similar over most of the focal range.

The similarities between the shapes of the translation graphs from separate calibrations indicate the potential for a high degree of repeatability, and a calibration that is partly stable over time. However, as shown by the y-translation graphs, use of the camera can introduce a global translation error into the calibration. This is probably due to the disturbance of one or more elements within one or both of the lens assemblies.

6.3.10 Summary of the errors in the lens calibration

The standard errors for the lens calibration derived in section 6.3.4 and section 6.3.6 are summarised in Table 10:

camera	Δx/pixels	Δy/pixels	Δc/c
left	0.124	0.142	0.000129
right	0.190	0.220	0.000168

Table 10 - summary of lens calibration random errors

These are independent errors, so combine using RSS (root sum squares), to give the total lens calibration errors for the camera pair. Converting all errors to pixels and truncating to two decimal places produces the results in Table 11:

	Δx/pixels	Δy/pixels
translation error/pixels	0.23	0.26
zoom error/pixels	0.14	0.11
total random error/pixels	0.27	0.28

Table 11 - total random errors from lens calibration

These errors are valid on short timescales (days, weeks), and are suitable for use in comparison with other calibration methods. For longer timescales, an overall translation may be introduced due to shifts in the lenses, as described in section 6.3.9, the magnitude of which will depend on how the lenses are treated. For a month's general use, this shift was within 1.5 pixels.

The total errors shown above are dominated by the errors due to repositioning the lenses. These arise through incorrect positioning of the focus barrel (from human error in setting the rotation position, and from slack in the focus barrel thread), and through motion of all or part of the lens assembly due to forces applied during the positioning. Errors arising from the software calibration are estimated in section 6.1.8 to contribute only ±0.03 pixels.

6.3.11 Conclusions on lens calibration

Lens calibration provides an effective way of estimating both the focal positions necessary to match the scales of the images taken with a stereo camera, as well as estimating the translation between the images (and hence the cropping required).

The high linearity of the image scale with focus position implies that a simpler calibration could suffice to match image scale between cameras, and the stability over time is high enough that lenses are unlikely to need to be recalibrated for scale matching unless a larger than normal disturbance to the lens occurs. The difference between the two calibrations implies a maximum vertical disparity of less than 0.3 pixels at any point in the image, if the other calibration was used to match the scales at either time. The lenses are also very close in terms of their focusing threads, in other words a rotation of 1° on one lens's focus ring is very nearly equal in terms of scaling effect to 1° on the other's. The calibration for scale might therefore only need to be as simple as a single image with each camera to determine the offset of the focusing positions, in order to provide a useful calibration.

The variability in the translations between the two cameras' images over long periods of time means that one initial calibration is not sufficient to predict future translations with to an accuracy within one pixel. The similar shapes of the graphs suggest that an initial calibration to find the shape of the translation graph, along with single image pair measurements after moving the camera about to determine the graph's offset, might be enough to estimate translations with reasonable accuracy. The translation predictions can only be estimates because the lenses used are unstable systems (as described in section 3.4), and should be used as such rather than as a confident measure of the exact disparities.

If the lens system could be redesigned for greater stability, so that image translations were repeatable to within a pixel, then this calibration method could be used to predict the image translation, and necessary cropping, between a pair of stereoscopic images to sub-pixel accuracy. Kinematic mountings are designed for repeatability, and though not investigated during the course of this thesis, should be examined for suitability in future work (further details in section 10.5.1) to improve stereoscopic camera design in this respect.

7 The calibrated camera

7.1 Introduction

This chapter concerns the integration of the camera components into a complete system. The method for calibrating the sensor orientation has been laid out in chapter 5, but maintaining this orientation requires a mounting design which prevents subsequent movement of the components. 7.2 describes the mounting of the cameras, in terms of the mechanical components used and the method of assembly. 7.3 shows the completed camera system, while 7.4 illustrates some of the strengths of the camera system through sample images captured with it. 7.5 briefly describes some user reactions to stereoscopic images taken with the camera.

7.2 The camera mounting

7.2.1 Purpose and design principles

The purpose of the camera mounting is to hold the two cameras in their aligned relative positions. It needs to be rigid, to prevent the cameras from moving once mounted, and it needs to be designed so that the camera can be mounted at a range of orientations. The mounting should ideally make use of the screw holes in the camera body already present.

The technique used to mount the cameras is called potting. This is where the mount has two halves, one of which has pins which float inside the pots of the other. When the camera orientation is as desired, adhesive is poured into the pots while the half with the pins is held in position. The adhesive sets, and the two halves of the mount become fixed. To maximise the rigidity of the bond, the pins should be short and wide so that they do not allow much bending, and the width of the adhesive bond should be as small as possible, because glues are usually less rigid than the mount material.

The mounting should not only be rigid itself, but should hold the camera firmly, so that the camera cannot move within the aligned mounting. Three-point mounting techniques should also be used where possible to minimise stresses in the system, which could potentially give rise to differential motion of the components as the system relaxes.

7.2.2 Components

The camera mounting consists of three pieces. A base plate, shown in Figure 91, attaches to the base of the camera by two screws, and has three bolts protruding from the bottom of the plate which fit into holes in the rail plate, shown in Figure 92. These bolts are set tightly in place with adhesive so that they cannot move. The rail plate bolts to a linear translation stage[95] (Newport 443) aligned with the X-direction, which bolts on to the rail slider[92] (Newport PRC-3), which is in turn mounted and set in place with adhesive on the camera rail. Finally, the side plate, shown in Figure 93, attaches to the camera by a single screw and to the base plate by two screws, which prevents the camera from rocking back and forth on the base plate. These components hold the camera firmly in place, so that all that is required to align the camera and keep it there is to set the three bolts into the three corresponding holes in the rail plate with adhesive. Figure 94 shows the three mounting pieces combined.

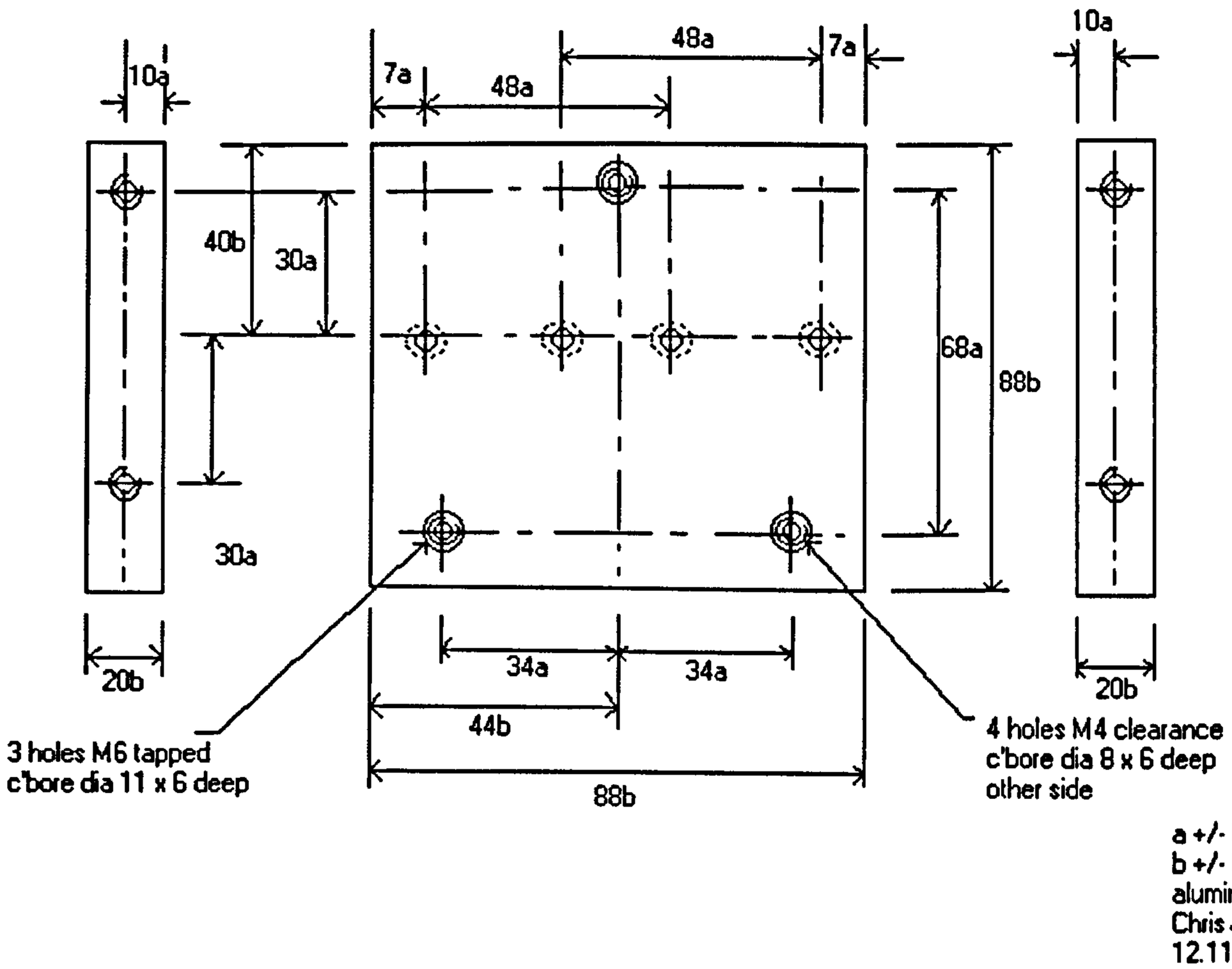

Figure 91 - base plate

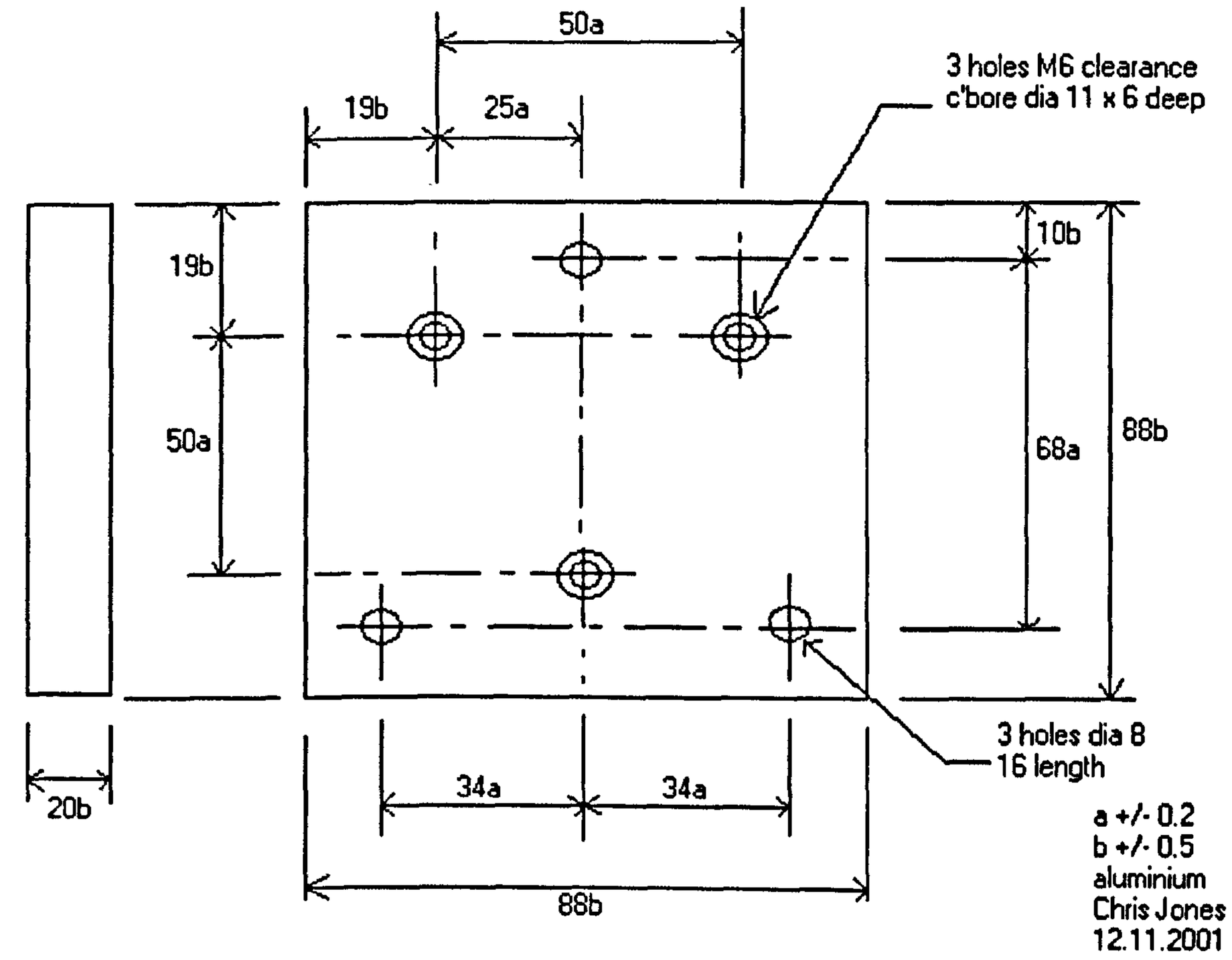

Figure 92 - rail plate

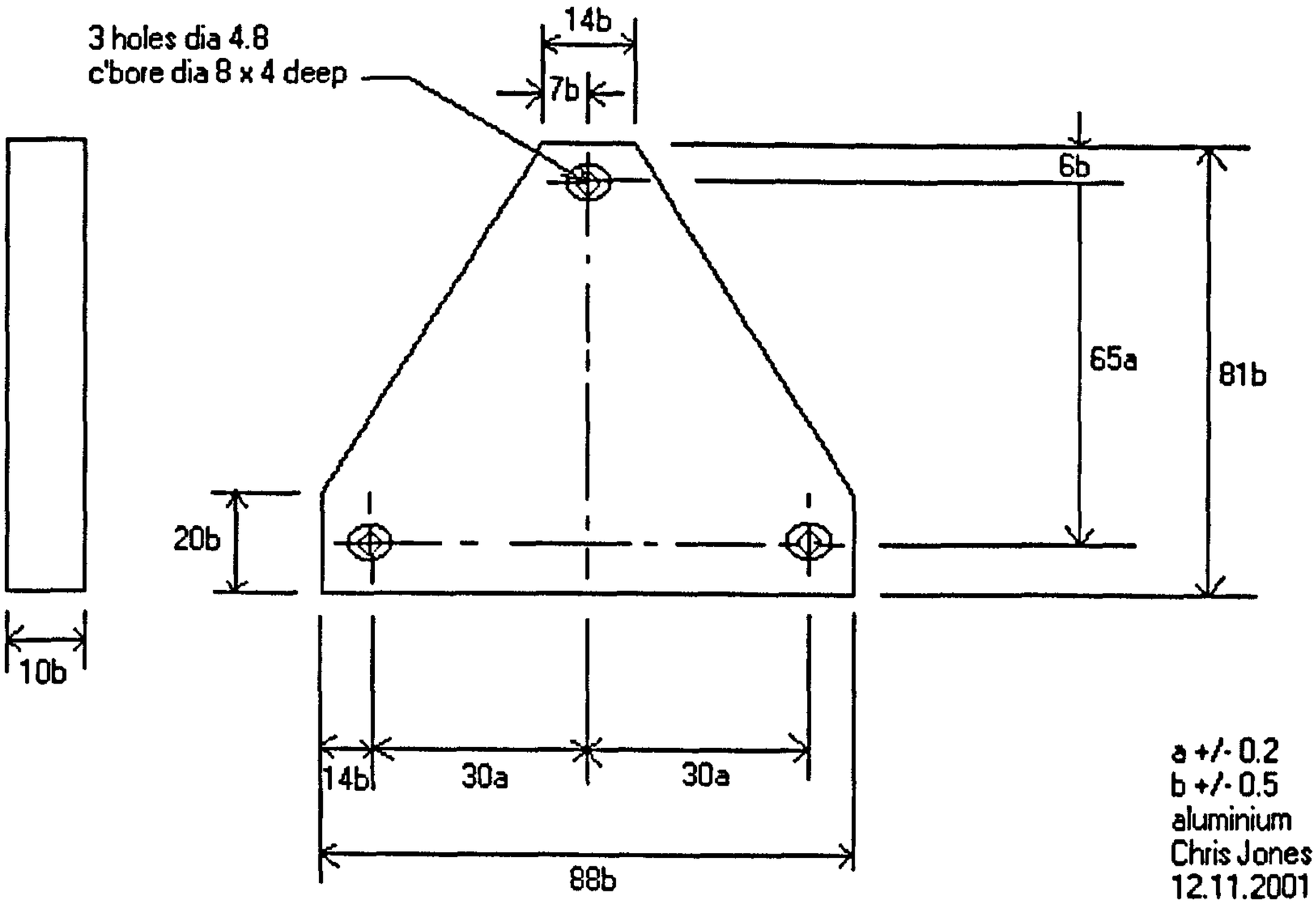

Figure 93 - side plate

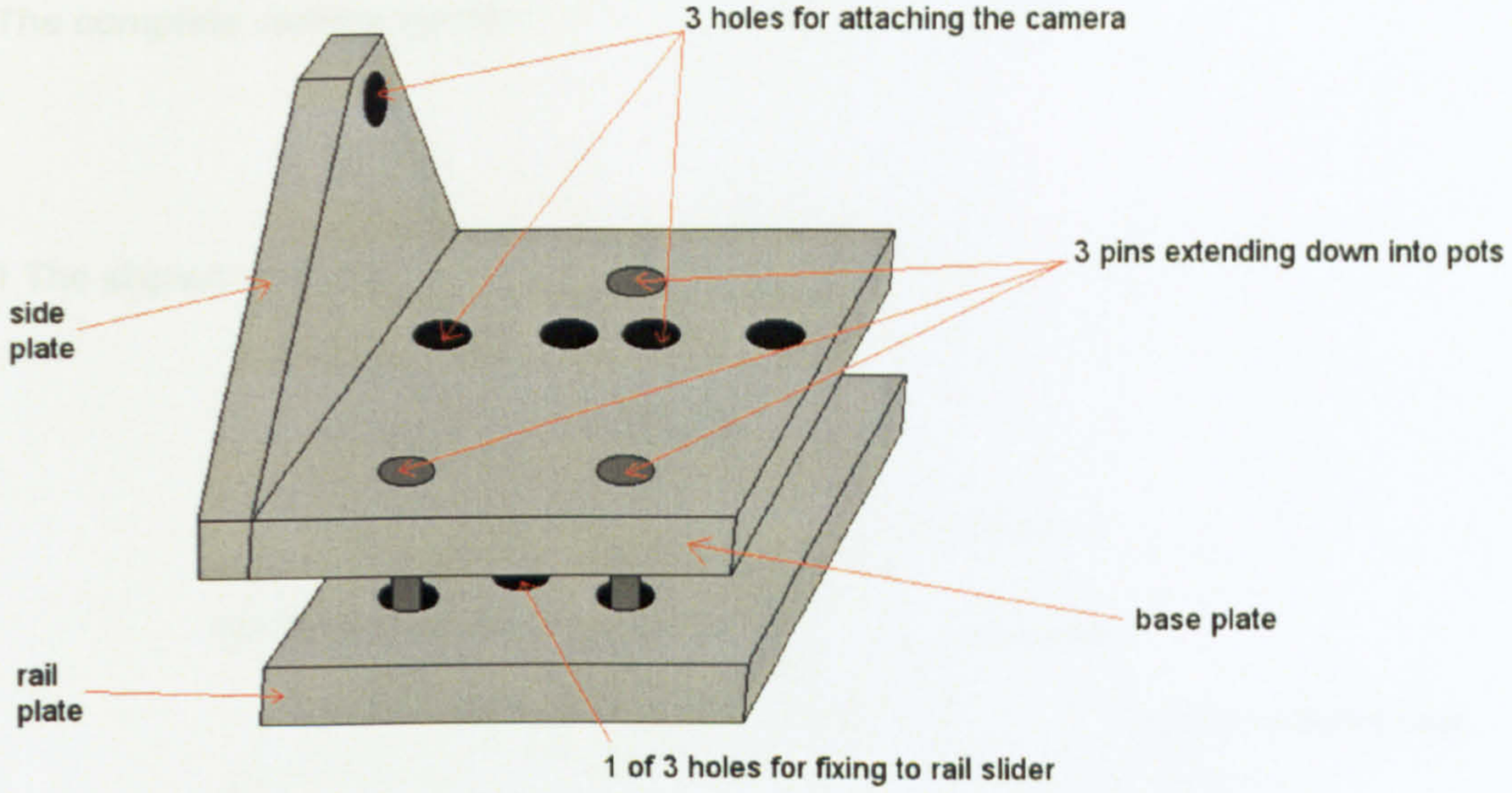

Figure 94 - the combined camera mounting, set up for the left camera

7.2.3 The adhesive used

The adhesive used is Scotchweld 1838[107], which is a two-part epoxy. It is used because of its very low outgassing (the release of gas from the adhesive over time), its high strength, and its medium length (24-hour) cure time. Shorter times make it hard to align the camera fast enough for the adhesive to remain liquid, while longer times make it difficult to construct a camera in a reasonable amount of time. The surfaces to be set together are thoroughly cleaned using iso-propyl alcohol before application, to prevent contamination from weakening the bond. Other adhesives tested include an optical lens bond (Summers Optical Inc. RD3-74), which hardened very rapidly initially, within 90 minutes, but remained slightly soft after 24 hours. The adhesive fractured under stress applied with the hands, and for this reason it shows itself to be less suitable than the 1838, which was resistant to such stresses. Other adhesives could be used, but they would have to meet the requirements of a convenient cure time and high strength combined with low to zero outgassing.

Low outgassing is of importance for several reasons, the most significant being that it induces low dimensional change in the bond over time. Excessive outgassing may both stress and weaken the bond, as well as releasing vapours which may potentially harm sensitive optical coatings.

7.3 The complete camera system

7.3.1 The aligned cameras

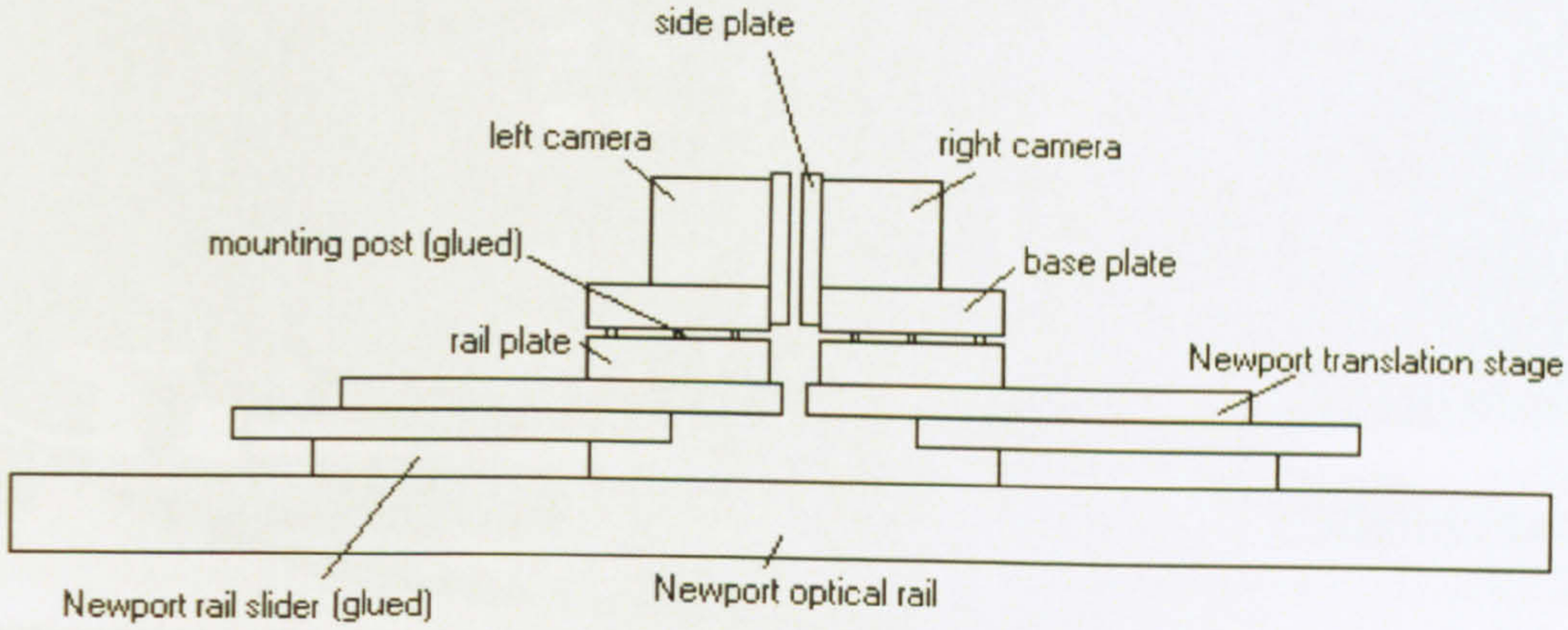

Figure 95 – diagram of camera setup, viewed from behind, minimum separation

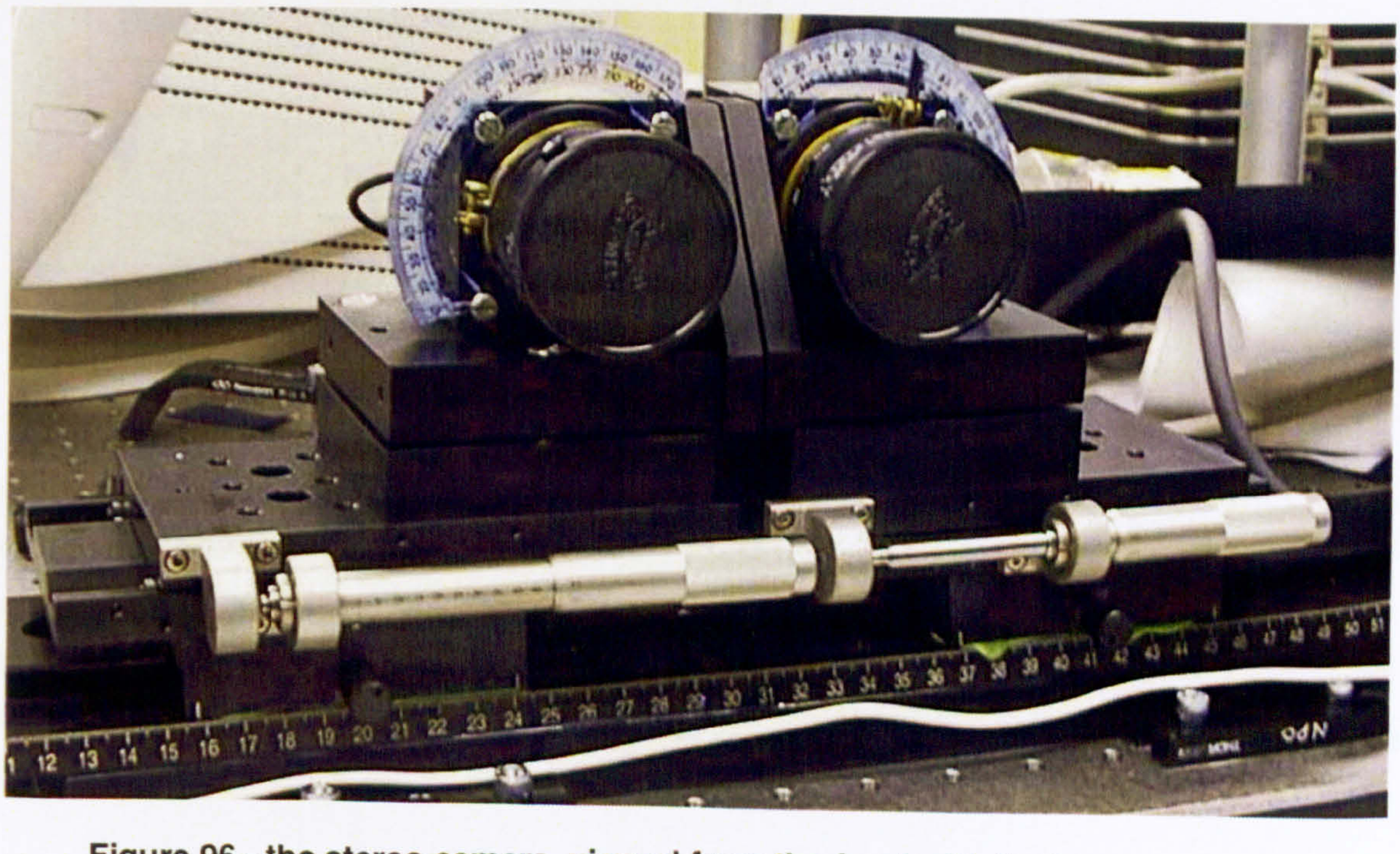

Figure 96 - the stereo camera, viewed from the front with the lenses attached

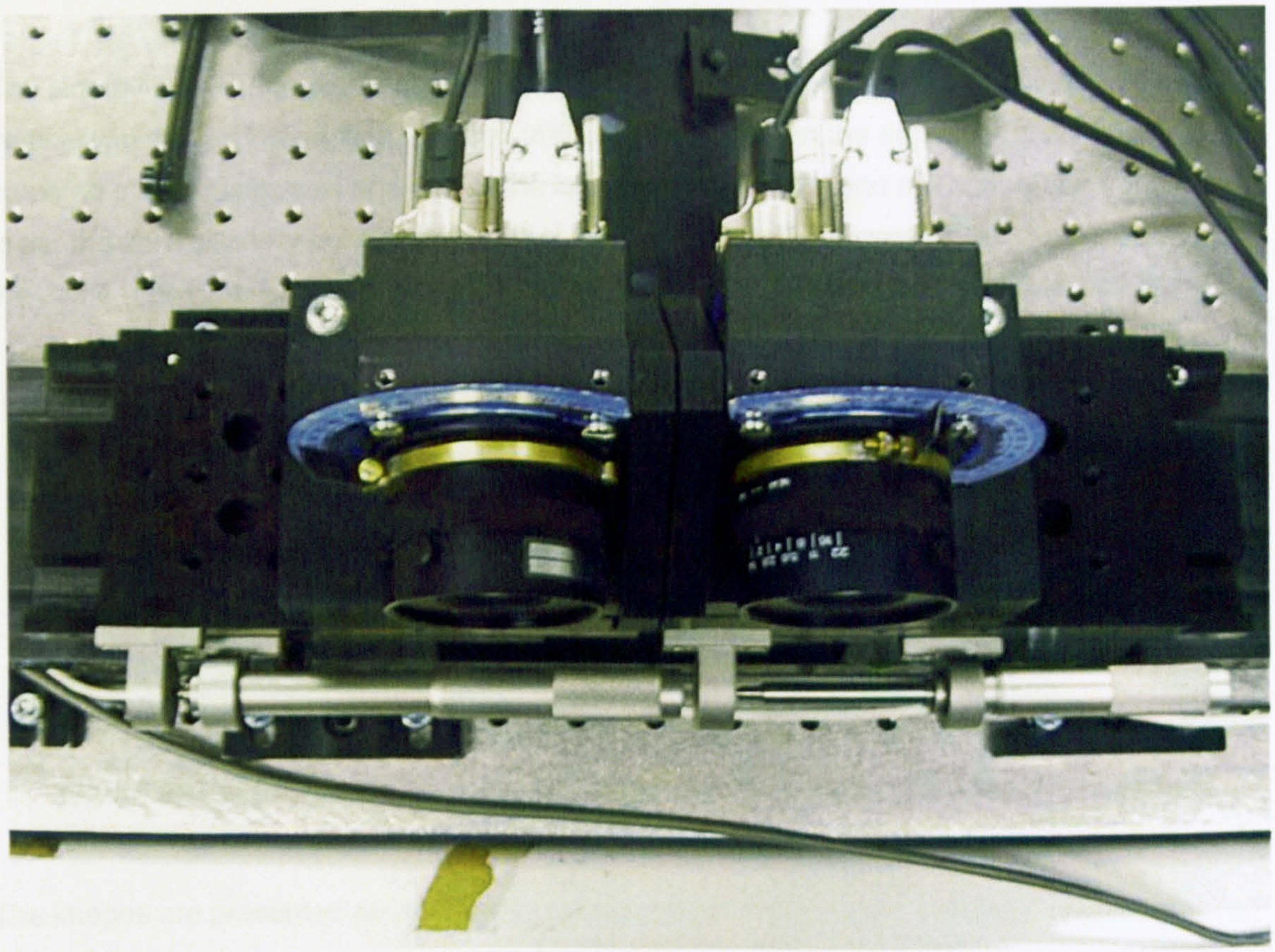

Figure 97 - the stereo camera, viewed from above, with the lenses attached

7.4 Sample stereoscopic images

Presented in this section are three examples of scenes imaged using the stereoscopic camera constructed in this thesis, using the methods described. The first illustrates close-range photography, the second shows a wider range of depths and the appearance of fast moving objects, and the third shows a more distant scene. Anaglyphs and side-by-side stereo are used to present the photographs, both methods have their shortcomings but presentation is limited by the thesis' format.

7.4.1 Close range photography - Schneider lens

A stereoscopic image pair was taken at close range, with the left lens focused at 83°. The right lens was set to 77.6° in accordance with the calibration given in Equation 48, and Figure 86 predicts a -6.3 pixel vertical disparity for these positions. The aperture used was f/3, and the exposure time was 20ms.

The side by side image pair is shown in Figure 98, having been cropped at 506 and 906 pixels vertically in the left image, and 500 and 900 pixels vertically in the right image. A horizontal cropping of 530 pixels was applied (at the left of the left image, and the right of the right image), to set the zero disparity point at the front of the lens being photographed, at around 23cm from the camera pair. No processing apart from the basic CFA interpolation has taken place.

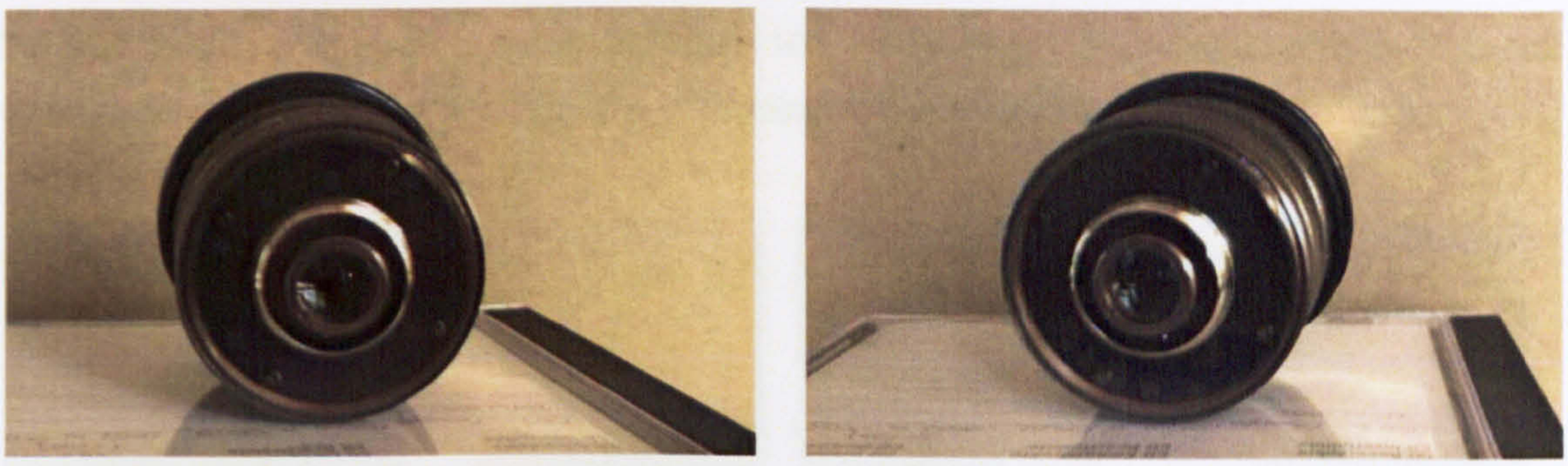

Figure 98 - side by side stereo photograph of a Schneider Cinegon 1.4/8mm lens

The images are presented as a red-green anaglyph for stereoscopic viewing in Figure 99:

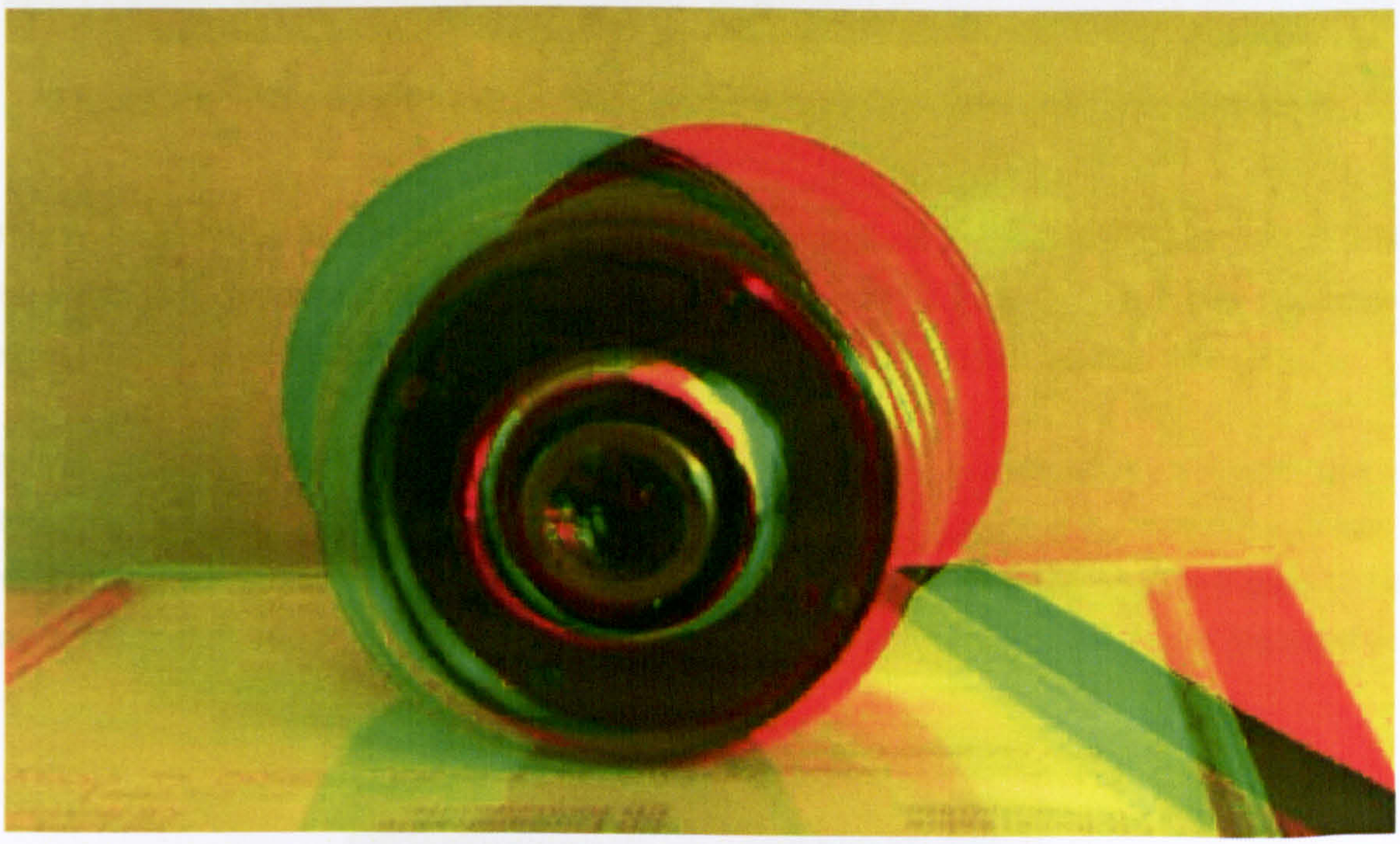

Figure 99 - red-green anaglyph stereo photograph of a Schneider Cinegon 1.4/8mm lens

This anaglyph is not an especially interesting stereo photograph, but it illustrates the good vertical alignment even at close range.

7.4.2 Motion - screwdriver and desk

A stereoscopic image pair was taken of a desk with a falling screwdriver to provide motion, with the left lens focused at 69°. The right lens was set to 63.5° in accordance with the calibration given in Equation 48, and Figure 86 predicts a -7.0 pixel vertical disparity for these positions. The aperture used was f/3, and the exposure time was 70ms.

The side by side image pair is shown in Figure 100, having been cropped vertically by 7 pixels. No processing apart from the basic CFA interpolation has taken place.

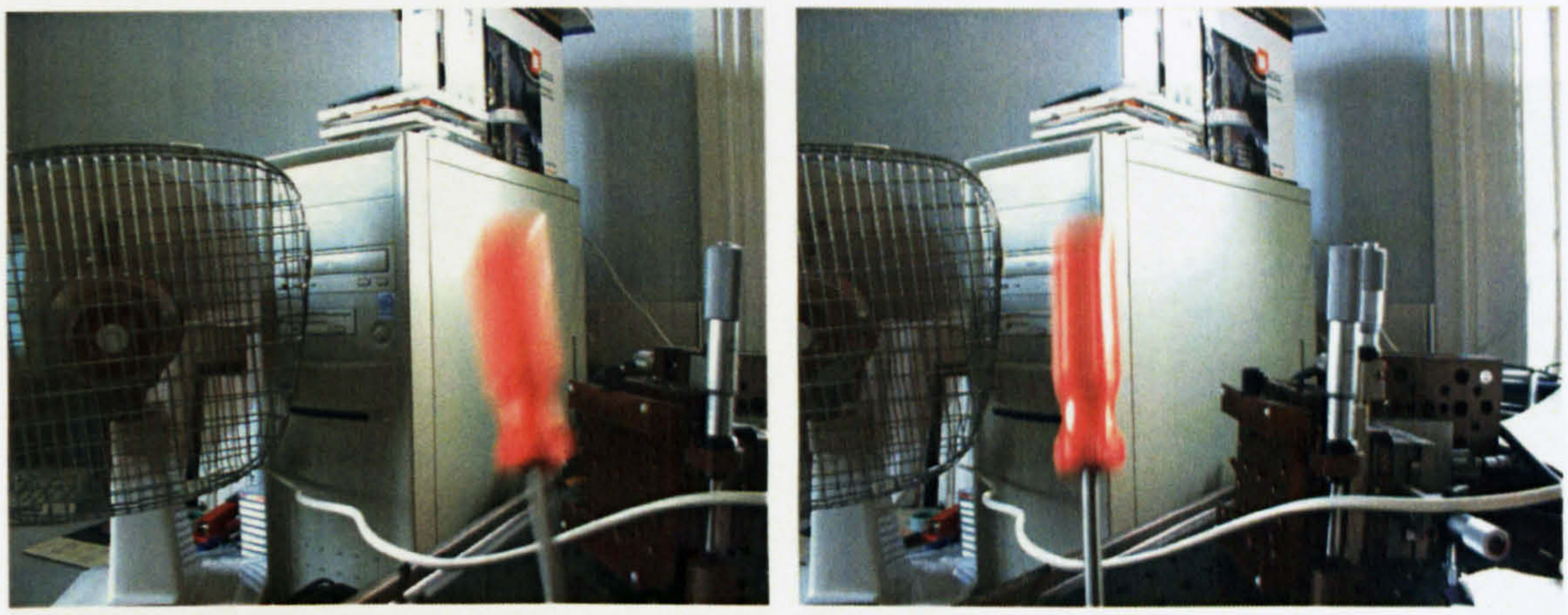

Figure 100 - side by side stereo photograph of a desk scene with a moving object

This image pair illustrates the importance of synchronisation, as discussed earlier in 4.1.18. Had the error in synchronisation been larger, the moving screwdriver would be in different locations when imaged by each camera head, and binocular fusion of the object would be damaged.

This image also clearly illustrates the validity of the lens calibration's cropping prediction. The floppy disk slot near the centre of the images in Figure 100 is enlarged and presented with the heights as predicted by the calibration, in Figure 101 below. Vertical cropping from Figure 100 is identical for each image, so that the vertical alignment is shown in accordance with predictions.

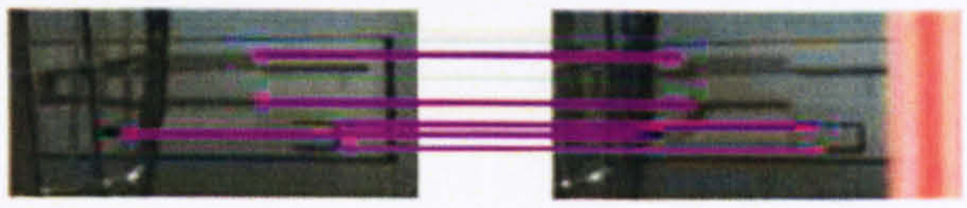

Figure 101 - enlarged portion of the stereo image to illustrate vertical alignment. The horizontal pink lines show image features vertically aligned to within one pixel

At this scale, individual pixels are clearly visible, and the images can be seen to be at the same height. The lens calibration's predictions are shown to be accurate.

7.4.3 Room scene

A stereoscopic image pair was taken of a room scene, with the left lens focused at 64°. The right lens was set to 58.5° in accordance with the calibration given in Equation 48, and Figure 86 predicts a -7.4 pixel vertical disparity for these positions. The aperture used was f/3, and the exposure time was 70ms.

The side by side image pair is shown in Figure 102, having been cropped vertically by 7 pixels. A horizontal cropping of 40 pixels has been used to place the zero disparity point. No processing apart from the basic CFA interpolation has taken place.

Figure 102 - side by side stereo photograph of a room

In this scene, the range of horizontal disparities is well suited to stereoscopic viewing, and a red-green anaglyph of the scene is presented in Figure 103:

Figure 103 - red-green anaglyph stereo photograph of a room scene

Again, vertical disparities are small and easily tolerable within the limits of the human visual system.

7.4.4 Outdoor scene

A stereoscopic image pair was taken of a room scene, with the left lens focused at 65°. The right lens was set to 59.1° in accordance with the calibration given in Equation 48, and Figure 86 predicts a -7.4 pixel vertical disparity for these positions. The aperture used was f/3, and the exposure time was 5ms.

The side by side image pair is shown in Figure 104, having been cropped vertically by 7 pixels. A horizontal cropping of 15 pixels has been used to set the zero disparity point. No processing apart from the basic CFA interpolation has taken place.

Figure 104 - side by side stereo photograph of an outdoor scene

In this scene, the range of horizontal disparities is mostly (with the exception of the railing) well suited to stereoscopic viewing, and a red-green anaglyph of the scene is presented in Figure 105:

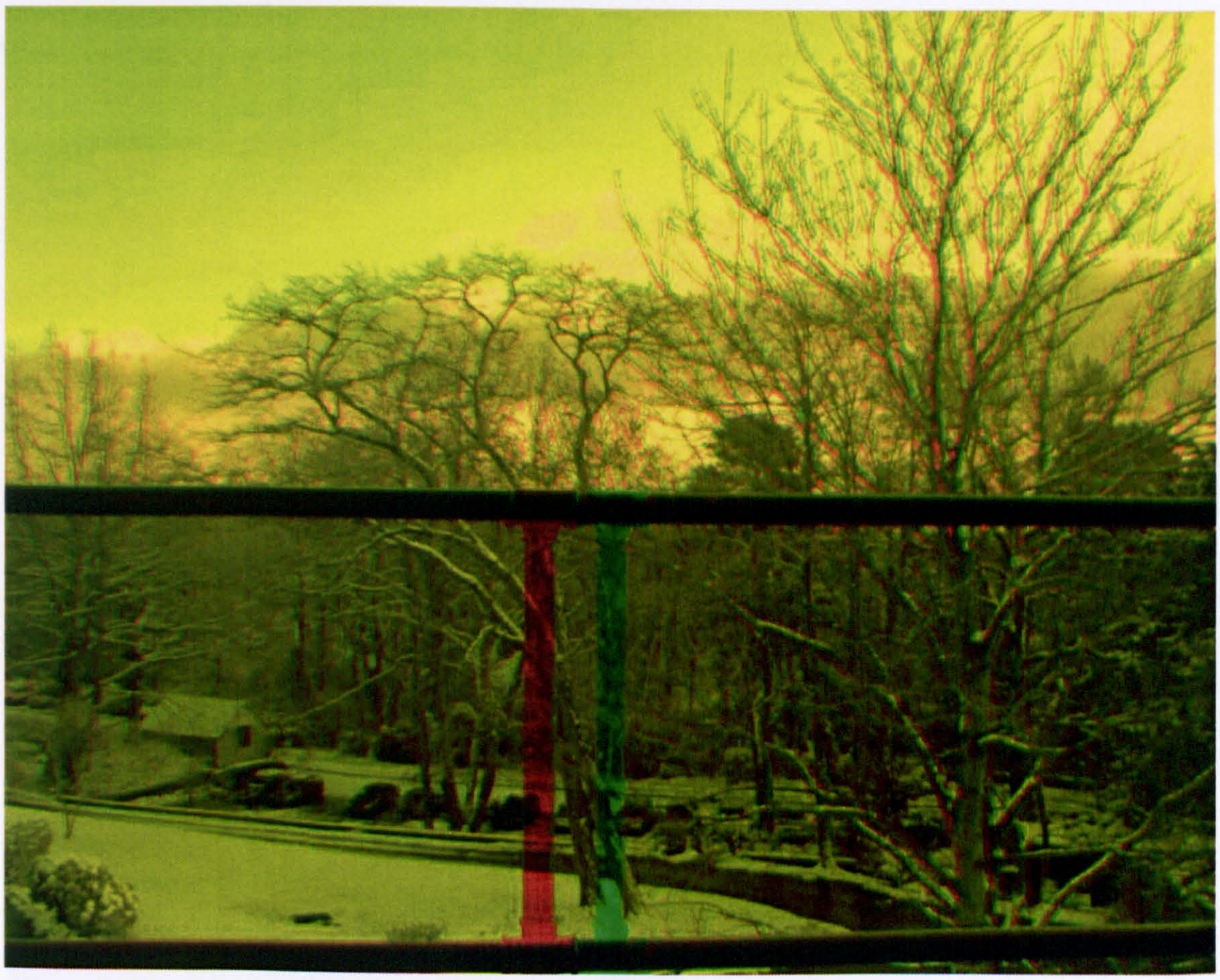

Figure 105 - red-green anaglyph stereo photograph of an outdoor scene

Measuring individual points in the scene shows how closely the vertical disparities match that predicted by the calibration of 7.4 pixels. Some measured points in the above image pair are shown below in Table 12:

Approximate position	Feature measured	Left image feature position	Right image feature position	Vertical disparity / pixels
Top left	-	-	-	-
Top centre	-	-	-	-
Top right	Branch joint	1245,36	1225,30	-6
Centre left	Branch joint	81,480	70,473	-7
Centre	Branch crossing	699,456	681,450	-6
Centre right	Branch joint	1267,438	1247,432	-6
Bottom left	Tree base	78,885	66,877	-8
Bottom centre	Footprint	532,977	517,970	-7
Bottom right	Snow clump	1214,911	1194,905	-6

Table 12 - features measured in the outdoor scene. No features were available for measurement in the top left and top centre areas. Features are measured by hand, and are subject to an error of approximately ±1 pixels

These vertical disparities are shown in their approximate locations in Figure 106 below:

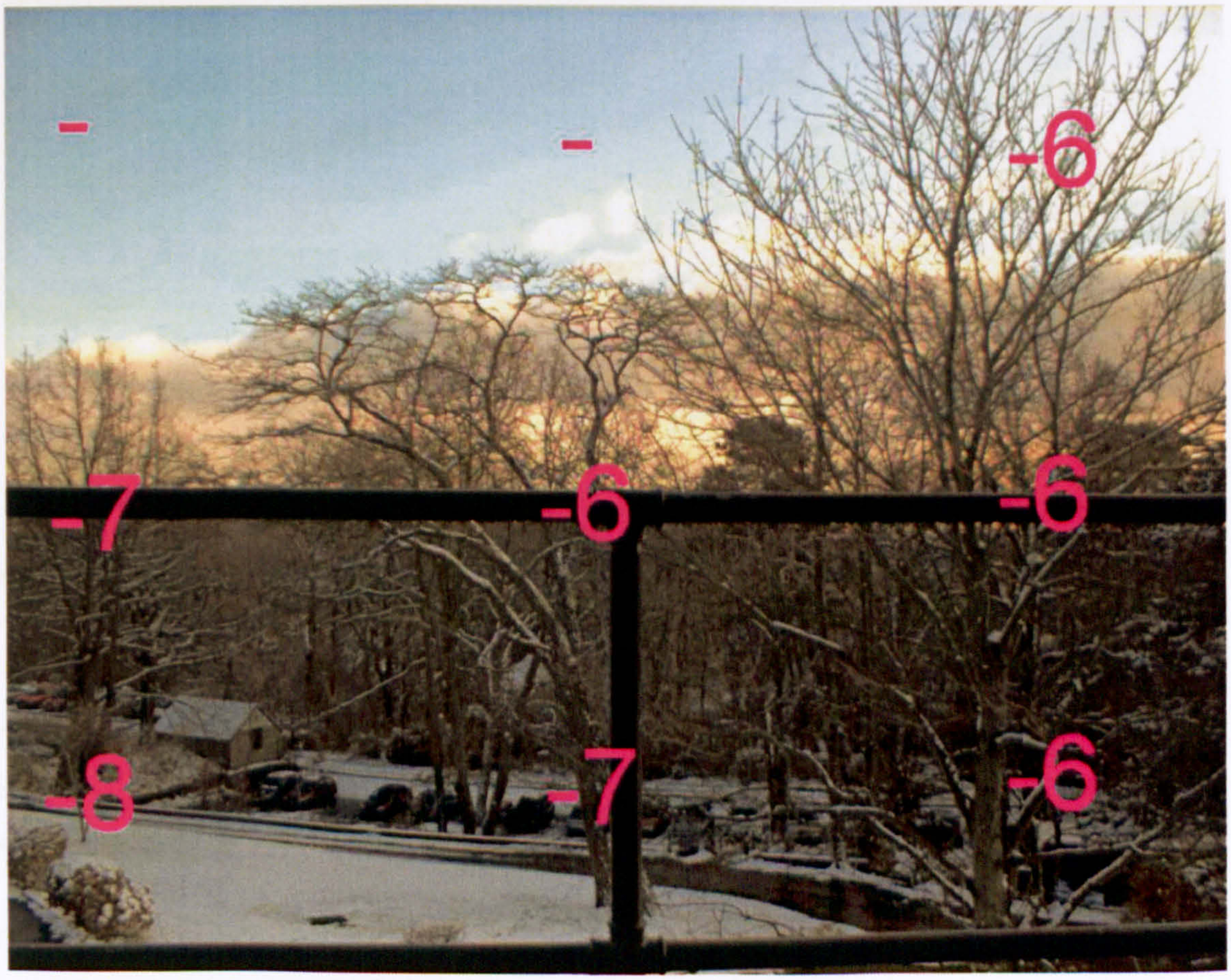

Figure 106 - vertical disparities in the stereo image of the outdoor scene, shown on the left camera image

These disparities are within ±1.4 pixels of the prediction, which is consistent with a point measurement error of ±1 pixels and the ±1.5 translation error described in section 6.3.11. The pattern of the disparities is suggestive of a small keystoning type error, which is investigated further in section 9.2.

7.5 User reactions

Many test images, including those above, have been shown on autostereoscopic displays for viewing by staff at Sharp Laboratories of Europe Ltd. and Sira Ltd. Feedback has been generally positive, with no problems comfortably fusing any parts of the images (where control over horizontal cropping is available). This indicates that the intended restriction of vertical disparity through calibration has been achieved.

Figure 107 - demonstration stereoscopic image pair taken at SLE

7.6 Summary

This chapter has given an overview of the completed camera system. The physical characteristics of the camera system have been described, and the image characteristics have also been demonstrated through the use of a range of example images taken with the camera.

Though this chapter shows that errors appear small to the human eye, it is also possible to quantify the performance of the camera and its calibration, and to compare this with existing camera calibrations. This comparison follows in chapter 8.

8 Comparison with existing calibration methods

8.1 Introduction

This chapter compares the calibration described in this thesis with the best existing calibration schemes for which experimental errors have been published. 8.2 deals with errors in the principal point position and principal distance, setting the basis for comparing errors across different camera systems and summarising the relevant data from each publication. 8.3 defines the most meaningful ways to compare the calibrations, shows that comparison, and discusses the relative merits of the calibration. In 8.4, published values for the total pixel error caused by calibration errors are compared with the calibration described in this thesis.

Where error estimates are given for previously published calibration methods, the quoted errors refer to the camera systems used by the authors of those publications, and are the errors resulting from their own experiments.

8.2 Comparison of the techniques used with software correction methods

8.2.1 Basis for comparison

Software correction of stereoscopic image pairs is difficult to correctly achieve. Rotation errors can be compensated for by warping the images, but camera translation errors may be fundamentally uncorrectable. If a scene contains occlusions, the visible parts of the scene depend on the viewpoint (for a camera, this is the optical centre), so an object visible from the viewpoint of an ideal camera may not be visible from the actual camera's position. This kind of error cannot be properly compensated for, since the corrected image requires information not present in the initial images. Any correction for translations, even if unencumbered by occlusions, requires an implicit determination of the depth of every point in the images, as well as their stereoscopic correspondence, which is difficult to achieve accurately and reliably in an automated way[108]. Because of this, a hardware calibration for camera translation should be used.

Self-calibration, as outlined in section 2.2.7, uses multiple images of a target from different viewpoints to determine both the intrinsic and extrinsic camera parameters. Some self-

calibration schemes have been tested for accuracy, and the results can be used to compare with the results from the work in section 6. Where data is taken from external references with more than one set of results, the set with the smallest errors is chosen for the best comparison.

8.2.2 Triggs[55]

Triggs' method uses 5 or more different views of a planar calibration target to recover the scene geometry and camera motion through a nonlinear minimisation, under the assumption that the internal parameters of the camera are constant. The target features are spots.

The data from "Autocalibration from planar scenes"[55] is given in Table 13:

	c	x_0	y_0
mean/pixels	1515	271	264
standard deviation/pixels	4	3	4

Table 13 - sample means and standard deviations of principal distance and principal point, for the calibration set[55], all values in pixels

Triggs notes that values of c seem suspiciously high and may have been subject to a small systematic error. The calibration values are the averaged results of several single-image calibrations using all visible points on the grid (49 at most). The frame size used is 512x512 pixels.

8.2.3 Zhang [1][85]

Zhang [1]'s method uses 2 or more different views of a planar calibration target to recover an estimate of the scene geometry and camera motion through a closed form solution, under the assumption that the internal parameters of the camera are constant. This estimate is refined using a nonlinear maximum likelihood estimation. The target features extracted are corners, and radial distortion is modelled. The best results are obtained by using 5 images.

The data from "A flexible new technique for camera calibration" is given in Table 14:

	c_x	c_y	x_0	y_0
mean	832.50	832.53	303.96	206.59
standard deviation	1.41	1.41	0.71	0.66

Table 14 - results of calibrations[85], all values in pixels. Values taken are maximum likelihood estimates for the five-image sequence

The principal distance c is calculated in both the x and y directions, as c_x and c_y. The two values represent possible different pixel pitches in the x and y directions, though their closeness indicates that they are probably the same. The frame size used is 640x480 pixels, and the lens used has a focal length of 6mm.

8.2.4 Takahashi et al[109]

Takahashi et al's method uses 1 or more views of a planar, parallelogrammetric calibration target to recover the scene geometry and camera motion through a series of steps. Perspective parameters are recovered from the shape of the parallelogram using a nonlinear optimisation. Rotational parameters are determined from the parallelogram's shape and orientation. Translation parameters are estimated from the size and position of the parallelogram, using a least squares method. Lens distortion parameters are estimated by least squares fitting a distortion model to the distorted grid, to straighten lines of points.

The data from "A Camera Calibration Method Using Parallelogrammetric Grid Points" is given in Table 15 and Table 16:

Camera 1	c_x	c_y	x_0	y_0
mean	1017.118	-842.160	324.352	111.215
standard deviation	56.057	46.948	7.602	15.461

Table 15 - perspective parameters[109] for camera 1, all values in pixels

Camera 2	c_x	c_y	x_0	y_0
mean	673.830	-562.931	320.771	136.751
standard deviation	30.507	26.339	3.119	4.428

Table 16 - perspective parameters[109] for camera 2, all values in pixels

The calibrated parameters for both cameras will be used for comparison. Camera 2 uses a very high distortion, 3.6mm lens, compared to a more conventional, medium distortion 11.5mm lens used for Camera 1. Both image sets use a frame size of 640x480 pixels. The negative values in Table 15 and Table 16 are due to the coordinate system used.

8.2.5 Devy et al[110]

Devy et al's method uses multiple views of a known planar calibration target to recover the scene geometry and camera motion through a nonlinear minimisation technique, under the assumption that the internal parameters of the camera are constant. The calibration tile is moved between images, rather than the camera. The target features detected are line intersections.

The data from "Camera calibration from multiple views of a 2D object, using a non linear minimization method" is given in Table 17:

image number	maximum error/pixels
1	1.057
2	0.497
4	0.848
5	0.544
8	0.587
10	0.923
11	1.209
12	0.577
13	0.563
15	0.735

Table 17 - errors evaluated for experiment 5[110]

Devy et al quote the total pixel error between the point positions in the images, and the predicted point positions from the calibrated camera parameters, which includes errors due to all parameters calibrated, including lens distortion. The maximum errors are chosen for comparison rather than the mean errors, because the calculation of disparity for other data sets will use the equations from 4, which calculate the maximum rather than the mean disparity for any given parameter error. Devy et al do not specify the characteristics of the sensor or lens used in the experiments. The errors quoted exclude those for 5 of the 15 total images used, where the error was unacceptably high[110].

8.2.6 Lavest et al[57]

Lavest et al's method uses multiple different views of both a known and an unknown three-dimensional calibration target to recover the scene geometry and camera motion through a

nonlinear minimisation, under the assumption that the internal parameters of the camera are constant. The target features extracted are circles.

The data from "Do We Really Need an Accurate Calibration Pattern to Achieve a Reliable Camera Calibration?" is given in Table 18, Table 19 and Table 20:

traditional	c_x	c_y	x_0	y_0
mean	1672.89	1676.03	386.74	276.86
standard deviation	0.208	0.209	0.426	0.365

Table 18 - calibration results with real data, traditional calibration, all values are in pixels[57]

new 1	c_x	c_y	x_0	y_0
mean	1672.85	1675.96	386.48	277.26
standard deviation	0.133	0.134	0.268	0.234

Table 19 - calibration results with real data, first new calibration, all values are in pixels[57]

new 2	c_x	c_y	x_0	y_0
mean	459.70	460.44	365.97	299.96
standard deviation	0.265	0.267	0.226	0.249

Table 20 - calibration results with real data, second new calibration, all values are in pixels[57]

The calibrated parameters for all three sets will be used for comparison. The traditional and first new calibration use a 10mm lens, while the second new calibration uses a high-distortion 3.6mm lens. All three image sets use a frame size of 640x480 pixels.

8.2.7 Zhang [2] et al[74]

Zhang [2] et al's method uses multiple different views of two calibration targets at different distances to recover an estimate of the scene geometry and camera motion through a closed-form solution, under the assumption that the internal parameters of the camera are constant. This estimate is refined using a nonlinear minimisation technique. Results from the two calibration targets are combined to improve the generality of the result. The target features extracted are corners.

The data from "An Effective Technique for Calibrating a Binocular Stereo Through Projective Reconstruction Using Both a Calibration Object and the Environment" can be summarised as a

single figure, that for their most general determination of the fundamental matrix applied to both the scenes they examine. The average pixel error between the point positions in the images and the epipolar lines is 0.72 pixels. This error measure only represents vertical disparity, and no estimate of horizontal errors is given.

8.2.8 Jones (work from this thesis)

The method presented in this thesis uses a single (for the parameters being compared, one is all that is required) image of a planar calibration target, from a stereo camera pair already aligned in rotation and translation. Camera parameters are determined using a closed-form solution. The target features extracted are points. Lens distortion is modelled and the effects corrected for by a calculated offset to the results, using an estimate of the distortion magnitude from the manufacturer.

The data from sections 6.3.4 and 6.3.6, with scales expressed as the focal length in pixels are given in Table 21 and Table 22:

left camera	c	x_0	y_0
mean	1194.03	-	-
standard deviation	0.154030	0.124	0.142

Table 21 - scale and translation errors, left camera, all values in pixels

right camera	c	x_0	y_0
mean	1194.03	-	-
standard deviation	0.200597	0.190	0.220

Table 22 - scale and translation errors, right camera, all values in pixels

The parameters derived within this thesis are relative, and so the values for the mean and standard deviation of c are expressed in pixel terms by scaling the data from 6.3.4 and 6.3.6 by a factor of 1194.03, which is the lens focal length of 8mm expressed in pixels. The frame size used was 1284x1024 pixels. The principal point itself is not calculated, but the standard deviation of the image translations from the calibration have the equivalent effect of errors in the principal point and are quoted for comparison.

8.2.9 Parameters for comparison

Compiling the results in Table 13, Table 14, Table 15, Table 16, Table 18, Table 19, Table 20, Table 21 and Table 22 produces the initial comparison table, Table 23:

	$\|c_x\|$	$\|\Delta c_x\|$	$\|c_y\|$	$\|\Delta c_y\|$	$\|\Delta x_0\|$	$\|\Delta y_0\|$
Triggs	1515.00	4.00	1515.00	4.00	3.00	4.00
Zhang [1]	832.50	1.41	832.53	1.41	0.71	0.66
Takahashi et al 1	1017.118	56.06	842.16	46.95	7.60	15.64
Takahashi et al 2	673.83	30.51	562.93	26.34	3.12	4.43
Lavest et al old	1672.89	0.21	1676.03	0.21	0.43	0.37
Lavest et al new 1	1672.85	0.13	1675.96	0.13	0.27	0.23
Lavest et al new 2	459.70	0.27	460.44	0.27	0.23	0.25
Jones L	1194.03	0.15	1194.03	0.15	0.12	0.14
Jones R	1194.03	0.20	1194.03	0.20	0.19	0.22

Table 23 - comparison table for principal axis and principal point errors for several different methods

Values are all in pixels, and are absolute values.

8.3 Comparison of scale and translation errors

8.3.1 Parameters for direct comparison

Because the nature of the calibrations performed in sections 5 and 6 comprise a novel method, it is difficult to draw direct comparisons for all calibrated parameters. Triggs[55] sets out some quantities which can be used to compare calibrations, which include some that are applicable in this instance. They are the absolute value of the error in principal distance divided by the principal distance, and the absolute value of the error in principal point location (Δpp) divided by the principal distance.

The long-term durability of the calibration is not addressed, because none of the external references provide any data on this. Durability of the camera described in this thesis is summarised in section 6.3.9.

The parameters for comparison are calculated using Equation 50:

$$\left|\frac{\Delta c}{c}\right| = \frac{\frac{|\Delta c_x|}{|c_x|} + \frac{|\Delta c_y|}{|c_y|}}{2}$$

$$\left|\frac{\Delta pp}{c}\right| = \sqrt{\left(\frac{|\Delta x_0|}{|c_x|}\right)^2 + \left(\frac{|\Delta y_0|}{|c_y|}\right)^2}$$

Equation 50

Combining Equation 50 and Table 23 produces the error comparison table Table 24:

	$\lvert\Delta c/c\rvert/10^{-4}$	$\lvert\Delta pp/c\rvert/10^{-4}$
Triggs	26.4	33.0
Zhang	16.9	11.6
Takahashi et al 1	554.3	200.2
Takahashi et al 2	460.3	91.3
Lavest et al old	1.3	3.4
Lavest et al new 1	0.8	2.1
Lavest et al new 2	5.8	7.3
Jones L	1.3	1.6
Jones R	1.7	2.4

Table 24 - comparison of errors in principal distance and principal axis for several different methods

The values for the hardware calibration errors (Jones) can be seen to be among the lowest among the errors compared. Only the method used by Lavest et al gives smaller errors.
The errors introduced by uncertainties in the principal distance and principal point can also be expressed as pixel errors, using Equation 31, Equation 32 and Equation 33. The standard deviation of the total error for random errors in principal point and distance is the RSS (root sum square) of the errors due to principal point and distance:

$$|\Delta x| = \sqrt{|\Delta x_0|^2 + |x - x_0|^2 . \left|\frac{\Delta c}{c}\right|^2}$$

$$|\Delta y| = \sqrt{|\Delta y_0|^2 + |y - y_0|^2 . \left|\frac{\Delta c}{c}\right|^2}$$

Equation 51

Calculating the pixel errors using Equation 51 gives the pixel error comparison table for principal point and axis errors, Table 25:

	Δx	Δy
Triggs	3.08	4.06
Zhang	0.89	0.78
Takahashi et al 1	19.30	20.53
Takahashi et al 2	15.06	11.92
Lavest et al old	0.43	0.37
Lavest et al new 1	0.27	0.23
Lavest et al new 2	0.29	0.29
Jones L	0.15	0.16
Jones R	0.22	0.24

Table 25 - comparison of pixel errors caused by uncertainties in principal point and distance for several different methods

Values are all in pixels, and are absolute values. Again, the errors for the hardware calibration (Jones) are among the lowest of those compared, for the parameters under comparison.

8.3.2 Evaluation of comparison of pixel errors

Table 25 shows that the calibration method used in this thesis is of similar accuracy as the best other calibrations, based on the comparison presented. On one hand, the comparison favours the method used in this thesis because the method is tuned to the simple demands of the situation, where low distortion lenses are specified and the cameras are prealigned in rotation and translation. The other calibration methods attempt to recover the principal point and distance parameters, and include the effects of radial distortion, along with varying camera positions in space, which greatly complicate the situation, compared to the simplified approach taken in the thesis of analysing image scale and translations. On the other hand, the external calibration methods do not address the problems of refocusing the lens, and do not contend with such high image noise as that found in the calibration used in this thesis. The image resolutions are also smaller than used in this thesis, so the errors they report will be magnified if images are presented on an equal scale.

The relevant issue in this comparison is suitability. By only requiring scale and translation determination from a calibration, it is possible to construct an algorithm that is simpler and more accurate for those results than a general purpose algorithm such as autocalibration. The keys to being able to exploit this idea are the laser alignment procedure given in 5, which removes the

need to recover or vary the world coordinates X, Y and Z, and the use of low-distortion lenses to remove the need to recover distortion parameters.

8.4 Total image error due to errors in calibrated parameters

Combining the errors introduced by the laser alignment procedure given in section 5.3.9 with the errors introduced by the lens calibration procedure given in section 8.3 gives the total image disparity produced by errors in rotation, translation, slidebar rotation, principal point and principal axis. The combined values for the camera pair are shown below in Table 26:

Δx	Δy
0.60	0.49

Table 26 - total image error in x and y directions due to uncertainties in calibrated parameters, all values in pixels

These values can be compared with the total image errors quoted by Devy et al[110], summarised in Table 17. The mean error per camera from Table 17 is assumed to be the error on each of a pair of cameras used, and the errors quoted in Devy et al are assumed to be equally distributed in the horizontal and vertical directions. The errors per camera are combined (the total error per camera pair being the RSS of the individual camera errors) and summarised alongside the mean error per camera pair for this thesis (from Table 4 and Table 11) in Table 27:

	mean total vertical image error per camera pair
Devy et al	0.75
Zhang [2] et al	0.72
Jones	0.49

Table 27 - comparison of mean total vertical image error due to uncertainties in calibrated parameters, values in pixels

The calibration performed in this thesis produces a lower vertical error than the best results from the Devy at al calibration, and lower than the error produced by the Zhang [2] et al calibration. It must be noted that the method used by Devy et al includes the modelling of effects of first-order radial distortion, which is not addressed in the calibration used in this thesis because it is not fully correctable in hardware. This means that the total error quoted by Devy et al includes one more error than the nine included in the figure for the calibration here. The Zhang [2] et al calibration may also include allowances for such effects. Bearing this in mind, it is reasonable to say that the errors due to rotation, translation, slidebar rotation, principal point and principal axis

uncertainties are roughly equivalent across the "software" (where alignment errors are entirely detected through image calibrations) and "hardware" (where as many errors as possible are eliminated in the initial instrument setup) methods. Lens distortion errors in the hardware calibration are considered in chapter 9.

9 Lens distortion

9.1 Restrictions on treatment of lens distortion

Lens distortion demands a different treatment from the other parameters examined in 4 because of its nature. The distortions caused cannot be eliminated in a single-camera design unless zero-distortion optics are used, and the effects are not simply translations that can be removed in a cropping operation. For these reasons, lens distortion has not been calibrated out in the camera constructed in this thesis, and therefore gives rise to image errors in addition to those summarised in section 8.4.

One way to treat the distortion might be to impose a standard distortion correction, based upon a calibration procedure[49]. However, there is another possibility, that of using different camera configurations to minimise the total image disparity, which is explored below. The concept of using a camera yaw to offset the effects of radial lens distortion is presented, and results from work carried out for this thesis in modelling the disparities caused by camera parameters. This technique has not been found in previous literature searches, and is thought to be novel.

9.2 Disparities due to lens distortion

A simple model for low distortion lenses is given in Equation 52[86]:

$$r = r'.(1 + k.r^2)$$

Equation 52

where r is the undistorted image radius, r' is the distorted image radius, and k is the first radial distortion coefficient. The inverse function[111] is given in Equation 53:

$$r' = r.(1 - k.r^2)$$

Equation 53

For a stereoscopic pair of distorted images of a frontoparallel plane, the disparities are equivalent to those introduced by distorting an image about two horizontally displaced points, the separation being equal to the stereoscopic horizontal disparity, and comparing corresponding points in the distorted images.

x,y are the undistorted image coordinates, x_L, y_L are the distorted left image coordinates, x_R, y_R are the distorted right image coordinates, x_{L0}, y_{L0} are the coordinates of the left distortion centre, and x_{R0}, y_{R0} are the coordinates of the right distortion centre.

The left image undistorted and distorted radii are defined in Equation 54:

$$r_L = (x - x_{L0})^2 + (y - y_{L0})^2$$
$$r_L' = (x_L - x_{L0})^2 + (y_L - y_{L0})^2$$

Equation 54

As the left distortion centre and left undistorted and distorted image points are collinear it follows that:

$$\frac{x_L - x_{L0}}{y_L - y_{L0}} = \frac{x - x_{L0}}{y - y_{L0}}$$

Equation 55

Substituting Equation 54 into Equation 53 and using Equation 55 gives Equation 56:

$$x_L - x_{L0} = (x - x_{L0}).(1 - k.((x - x_{L0})^2 + (y - y_{L0})^2))$$
$$y_L - y_{L0} = (y - y_{L0}).(1 - k.((x - x_{L0})^2 + (y - y_{L0})^2))$$

Equation 56

The right image undistorted and distorted radii are defined in Equation 57:

$$r_R = (x - x_{R0})^2 + (y - y_{R0})^2$$
$$r_R' = (x_L - x_{R0})^2 + (y_L - y_{R0})^2$$

Equation 57

As the right distortion centre and right undistorted and distorted image points are collinear it follows that:

$$\frac{x_R - x_{R0}}{y_R - y_{R0}} = \frac{x - x_{R0}}{y - y_{R0}}$$

Equation 58

Substituting Equation 57 into Equation 53 and using Equation 58 gives Equation 59:

$$x_R - x_{R0} = (x - x_{R0}).(1 - k.((x - x_{R0})^2 + (y - y_{R0})^2))$$
$$y_R - y_{R0} = (y - y_{R0}).(1 - k.((x - x_{R0})^2 + (y - y_{R0})^2))$$

Equation 59

The undistorted image coordinates can be expressed in terms of the distorted left image coordinates, using Equation 52, to give Equation 60:

$$x - x_{L0} = (x_L - x_{L0}).(1 + k.((x_L - x_{L0})^2 + (y_L - y_{L0})^2))$$
$$y - y_{L0} = (y_L - y_{L0}).(1 + k.((x_L - x_{L0})^2 + (y_L - y_{L0})^2))$$

Equation 60

The difference between the images, is defined by the disparities in the horizontal and vertical directions, Δx and Δy, as in Equation 61:

$$\Delta x = x_R - x_L$$
$$\Delta y = y_R - y_L$$

Equation 61

Equation 59 gives x_R and y_R as a function of x and y, and Equation 60 gives x and y as a function of x_L and y_L, so substituting Equation 60 into Equation 59 gives x_R and y_R as a function of x_L and y_L. Substituting this function into Equation 61 gives Δx and Δy in terms of x_L and y_L, allowing the disparities to be calculated in terms of the left image coordinates.

The values are assumed for the parameters, for illustration, are given in Equation 62 below:

$$x_{L0} = 582.3$$
$$x_{R0} = 701.7$$
$$y_{L0} = y_{R0} = 512$$
$$k = 4.49x10^{-8}$$

Equation 62

where those values correspond to disparities from an object at a distance of 1m from the perspective centre of an 8mm lens, and a 1284x1024 sensor with pixel size 6.7μm, with a maximum 3% barrel distortion (these correspond to the cameras used in this thesis). All distances are in pixels, and k is expressed in pixels^{-2}.

This produces disparity from radial distortion as shown in Figure 108 and Figure 109 below:

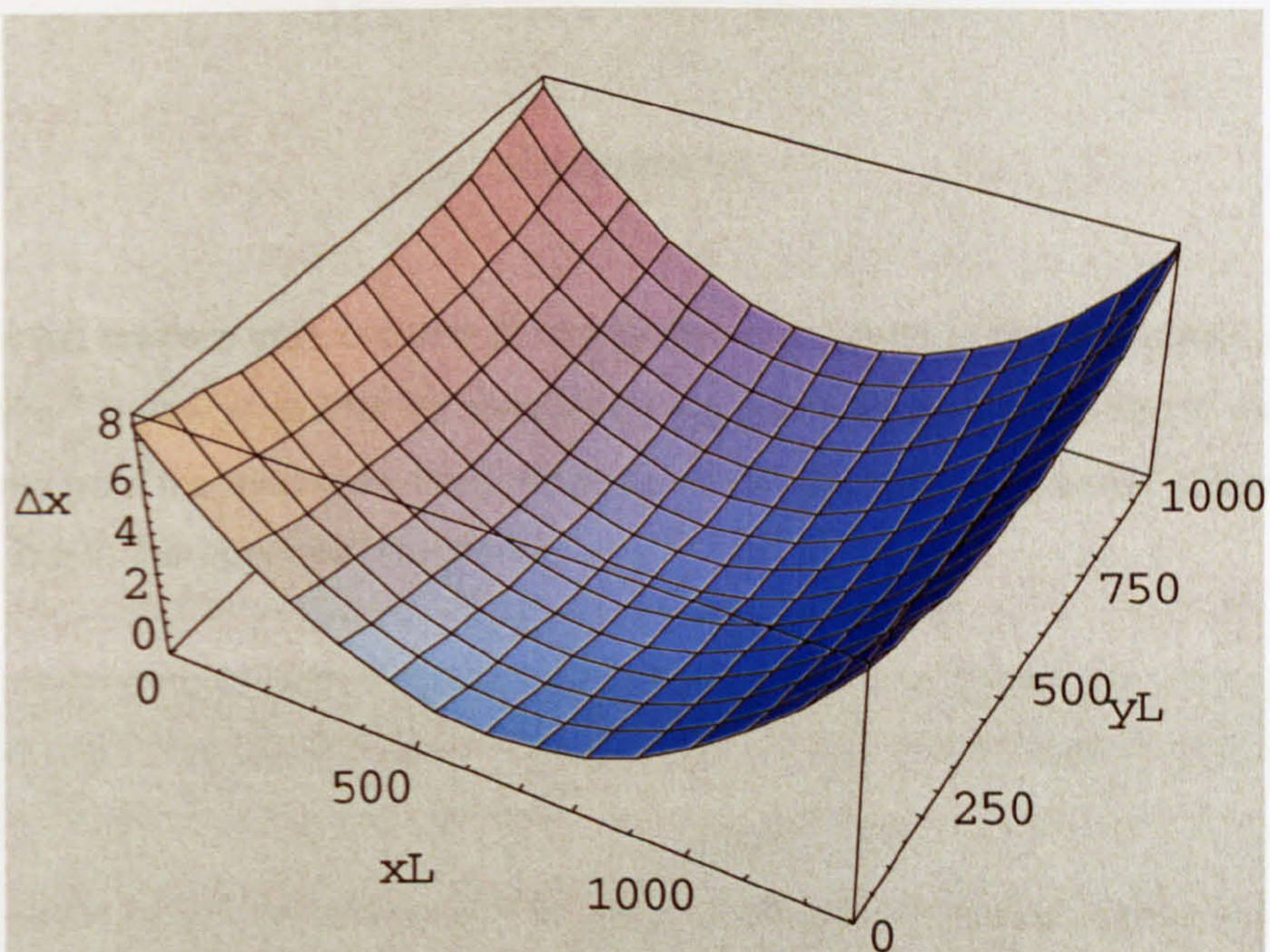

Figure 108 - horizontal disparity due to radial distortion

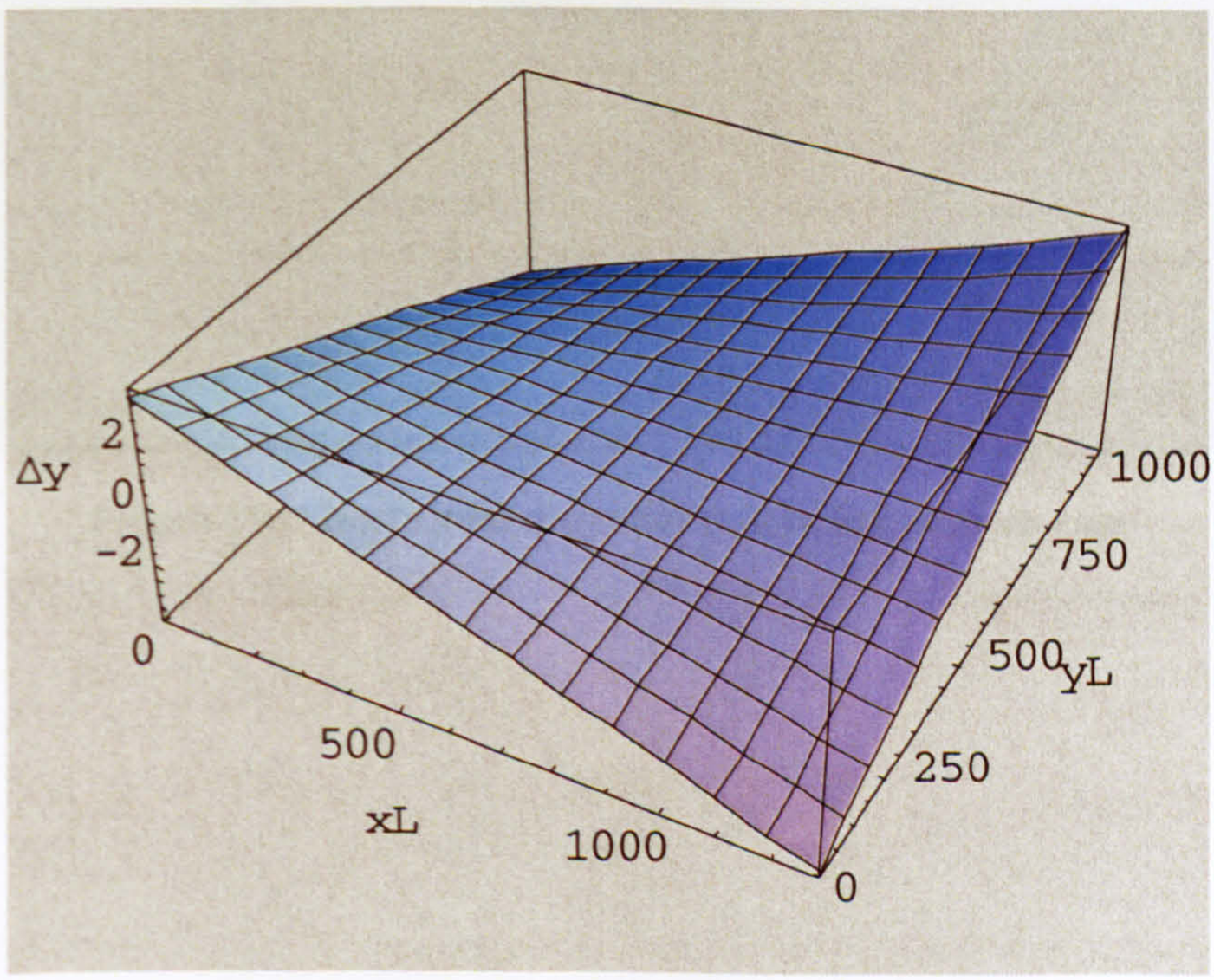

Figure 109 - vertical disparity due to radial distortion

The range of vertical disparity produced is between ±3.72 pixels.

From section 4.1.9, yaw introduces artificial disparities:

$$\Delta x = \frac{-\beta.(c^2.p^2 + (x_L - x_{L0})^2)}{c.p + \beta.(x_L - x_{L0})}$$

$$\Delta x = \frac{-\beta.(x_L - x_{L0}).(y_L - y_{L0})}{c.p + \beta.(x_L - x_{L0})}$$

Equation 63

Where β is the left camera yaw, c is the principal distance, and p is the pixel pitch. For c=8mm and p=149.3mm^{-1}, a yaw of -0.778° (a divergent yaw, numerically determined to cancel the vertical disparity from the lens distortion) produces a vertical disparity similar to that caused by the distortion, but in the opposite direction:

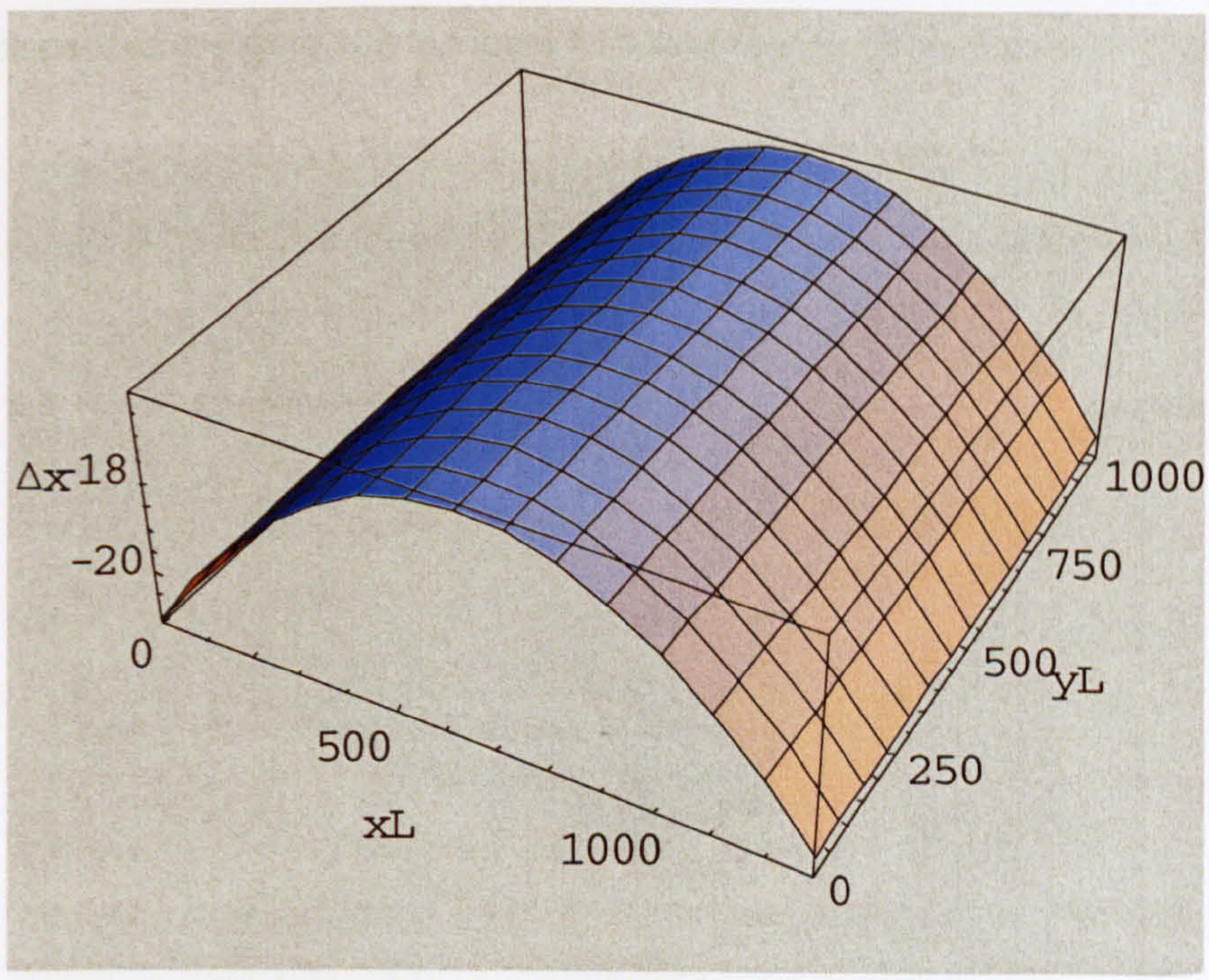

Figure 110 - horizontal disparity due to left camera yaw

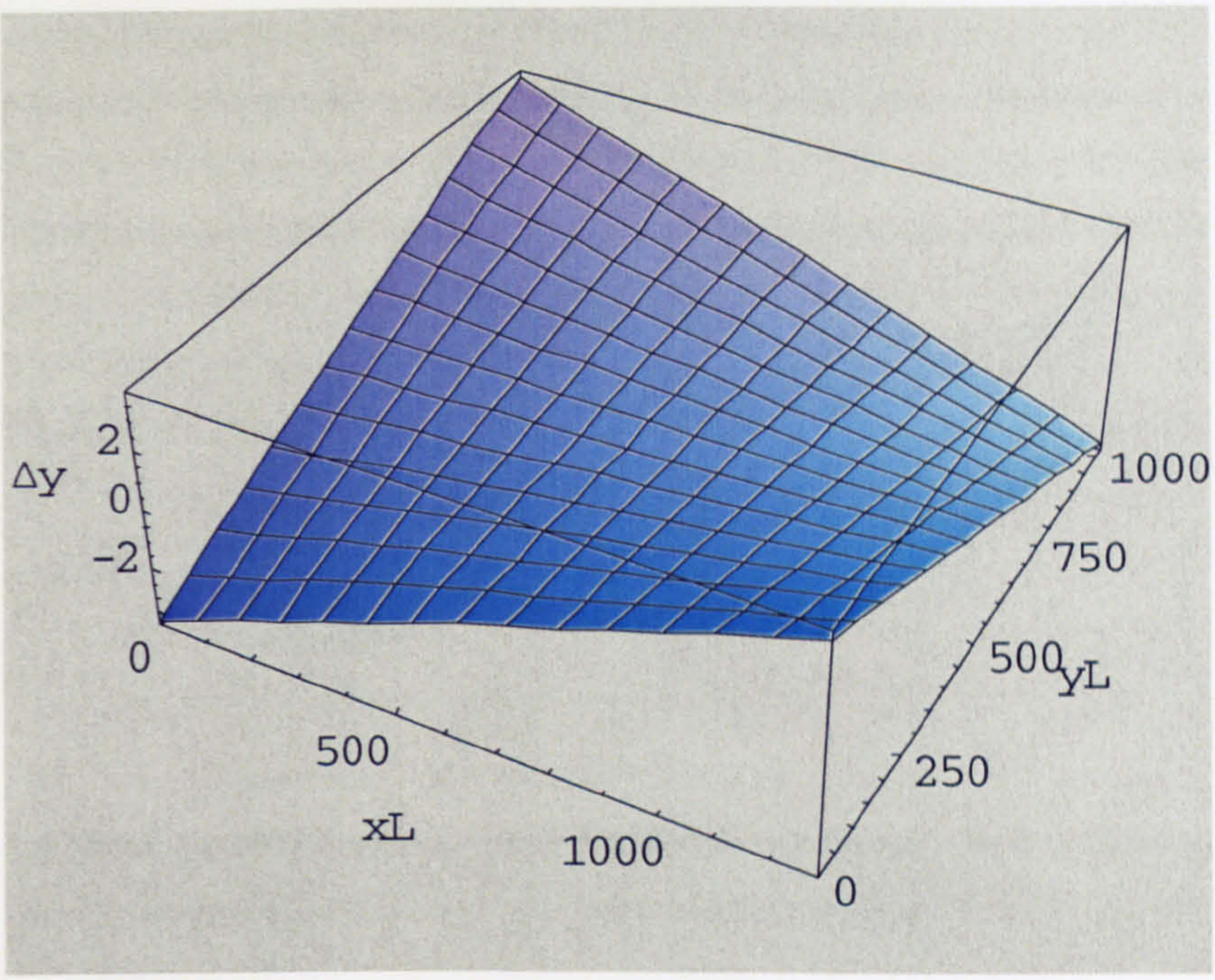

Figure 111 - vertical disparity due to left camera yaw

Adding the disparities in Figure 108 to Figure 110, and Figure 109 to Figure 111, gives:

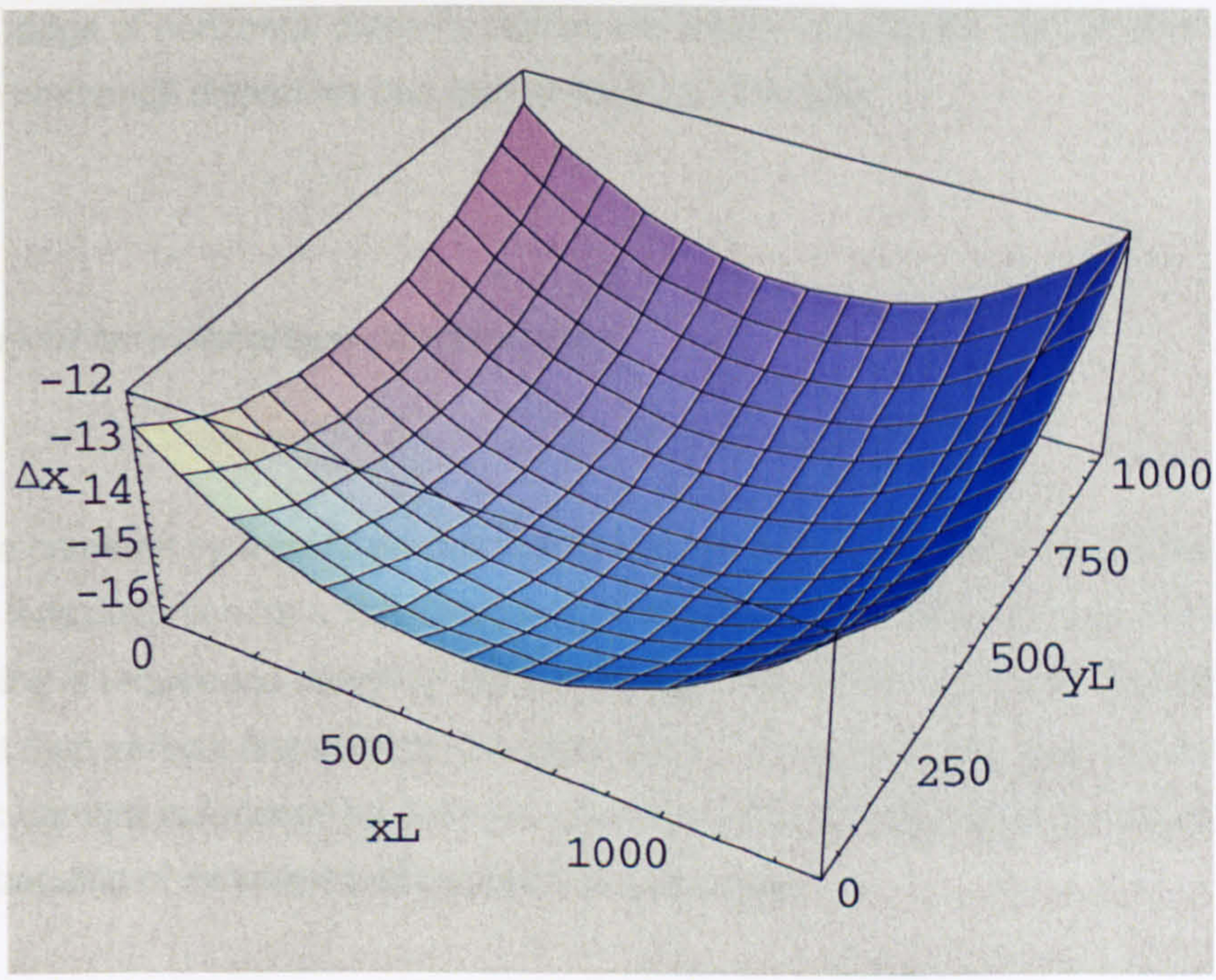

Figure 112 - horizontal disparity due to distortion modified by left camera yaw

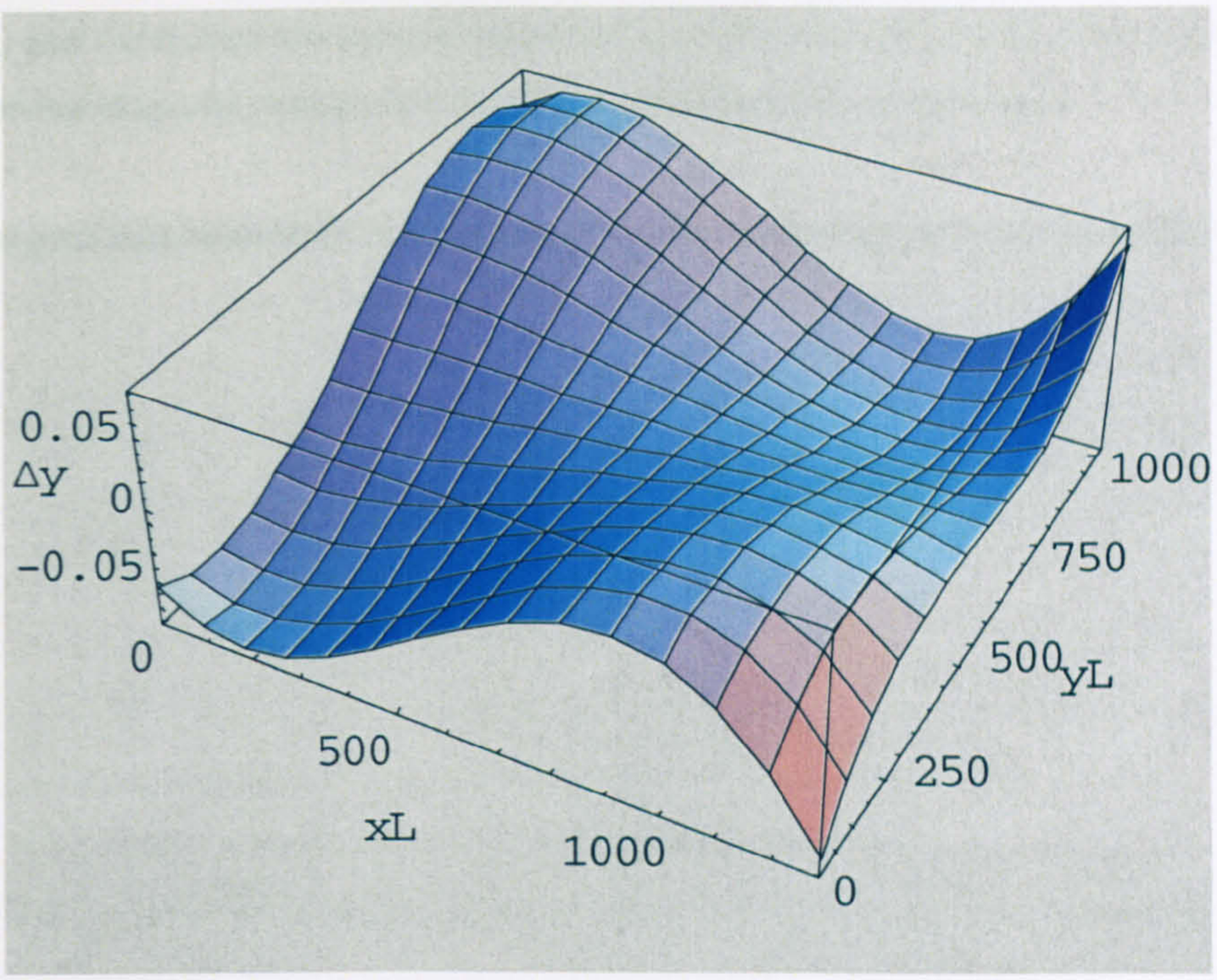

Figure 113 - vertical disparity due to distortion modified by left camera yaw

As Figure 113 shows, the vertical disparities caused by the distortion are reduced to within around 0.05 pixels across the image. Figure 112 shows a larger horizontal disparity than Figure 108, but the range of horizontal disparity across the image is reduced, and so with appropriate cropping the horizontal disparities can also reduced by the yaw.

9.3 Summary of lens distortion compensation

Diverging the cameras by a small amount could reduce the vertical disparity caused by simple barrel radial lens distortion for a specified horizontal disparity (or distance from the camera). Some cropping is required to minimise horizontal disparity, which could also be reduced (to a lesser extent than vertical disparity) by the procedure. The overlapping field of view of the stereoscopic camera is reduced by such an operation, and this disadvantage must be traded off against the benefits of minimising disparities from distortion.

Since vertical disparity due to lens distortion varies with horizontal disparity (determined by the subject distance from the camera), this compensation only achieves the highest vertical disparity reduction at the distance for which the yaw is calculated. A camera might be constructed to vary in yaw, so that the yaw could be changed to suit the scene being imaged. Another approach would be to set the yaw so that maximum vertical disparity is minimised over a range. For example, the maximum vertical disparity for zero yaw is ±3.72 pixels at 1m, and zero at infinity. Setting the divergent yaw to 0.386° removes half the vertical disparity (to ±1.86

pixels) at 1m, and increased the vertical disparity to ±1.86 pixels at infinity. This correction halves the vertical disparity range from 1m distance to the camera to infinity.

How any divergent yaw correction for radial lens distortion is applied must depend on the application.

10 Conclusions

At the beginning of this thesis in section 1.1, a number of objectives were defined. These objectives are restated in section 10.1 below, and the ways in which they have been met are described. An additional contribution of this thesis, the yaw correction for radial distortion, is also mentioned in 10.1.6.

Section 10.2 summarises the results of the camera calibration method as applied to the constructed camera, and this is discussed in section 10.3.

10.4 summarises the conclusions which can be drawn from the work contained in this thesis. 10.5 describes possible future extensions to the work, while 10.6 describes possible future applications of the work.

10.1 Key contributions

10.1.1 Relationship between human factors and digital images

Aim: To define the stereoscopic viewing requirements of the human visual system in terms of a digital image.

Based on the work contained in section 1.6, the relationship between the properties of the human visual system and digital stereoscopic images has been quantified.

10.1.2 Relationship between digital images and stereoscopic camera parameters

Aim: To determine the relationship between image point movement and camera component movement.

Based on the work contained in chapter 4, the relationship between image disparity and the spatial and temporal properties of a stereoscopic camera has been quantified. This has defined the basis for constructing stereoscopic cameras that produce images for comfortable human stereoscopic viewing.

Combined with the achievements described in 10.1.1, this defines the relationship between human stereoscopic viewing factors and stereoscopic camera calibration parameters.

10.1.3 Camera calibration methods

Aim: To devise a set of reliable camera calibration techniques which calibrate the camera system such that minimal correction of captured images is required, and for any correction to be calculated and non-interpolative.

The sensor alignment method described in chapter 5 forms a significant advance in the field of camera calibration. The method allows for a real time calibration of a digital sensor's rotation without requiring any image processing or complex targets. It also completely decouples the rotation and translation parameters of a camera for calibration, eliminating errors introduced by such coupling. This technique lends itself readily to simple automation, which would be an advantage if calibrating a large number of sensors.

The above sensor alignment method allows for a simple image-based lens calibration technique, described in chapter 6, which only has to calibrate principal point and distance parameters for a given focal position. The calibration technique allows for the use of a simple planar target, and uses a highly efficient and accurate closed-form technique to simply calculate the required parameters. The lack of coupling with rotational parameters further reduces the potential for errors in the calibration.

Calibrating the lens across its focal range also allows the calibration to remain valid when the lens is focused. This enables the capture of well-focused stereoscopic images with known vertical correspondence for scenes at any distance.

10.1.4 Use of advanced mounting techniques

Aim: To design and construct a stereoscopic camera system within a mounting framework which maintains the camera calibration while allowing flexibility of use.

The design and use of a mounting system based around the principle of three point mounting, described in 7.2, provides a strong, stable platform for the stereoscopic camera, with a low potential for motion due to mechanical stress. The combination of this with a potting technique using a strong and stable adhesive, and the alignment apparatus, allows a stereoscopic camera to be calibrated in hardware, and for this calibration to be highly durable.

10.1.5 Achieving comparable overall accuracy to that achieved using existing interpolative methods

Aim: To meet or exceed the standards of camera calibration available with existing interpolative methods.

The stereoscopic camera, constructed according to the calibrations in this thesis and described in chapter 7, has total image errors (summarised in section 8.4) due to rotation, translation, principal point position and principal distance of 0.60 pixels in the x direction, and 0.49 pixels in the y direction. This is a clear improvement on past work in the field.

10.1.6 Method of using a camera yaw to offset radial lens distortion effects

The method of offsetting radial lens distortion with stereoscopic camera yaw, as described in 9, allows the effects of such lens distortion to be reduced without requiring any interpolative correction or image post-processing. The yaw angle for a given stereoscopic camera may be varied in order to minimise vertical disparity for any scene distance from the camera.

10.2 Summary of results

Without any correction for lens distortion, vertical disparity in the calibrated camera's images is dominated by the lens distortion error. The total vertical error in this case is the combination of the errors from the alignment procedure, the lens calibration and the lens distortion, and is ±4.21 pixels at a scene distance of 1m from the camera. This is more than the ±3.4 pixels permitted by the size of Panum's area, under the conditions described in section 1.6.6. At scene distances of 1.3m or greater, the vertical disparity is less than ±3.4 pixels, and images taken will be suitable for viewing under the conditions described in section 1.6.6. This treatment is the one that has been applied to the camera described in this thesis.

10.3 Discussion of results

For the equipment detailed in section 4.1.3, the calibration methods outlined in this thesis can produce a camera to take stereoscopic images with a total vertical disparity within that tolerated by the human visual system (of ±3.4 pixels for a suggested display) for scenes at Z-distances of 1.3m or greater. The only software intervention required for this is a cropping, the amount of which is predicted by the calibration.

Most of the disparity is caused by differential lens distortion between the two cameras, which could be reduced by using lower distortion lenses, or by having a smaller horizontal camera separation. The lens distortion effect could be simply alleviated either in software or hardware, to reduce the vertical disparity to less than one pixel.

The method is therefore adequate for stereoscopic image capture for comfortable human viewing, as long as the objects to be imaged are further than 1.3m from the camera. Images taken with the camera pair (examples are given in section 7.4) verify the quality of the alignment and calibration by the small range of disparities exhibited.

The effort involved in setting up the camera hardware is considerable, but many of the required steps lend themselves to automation. The measurements required at each stage are simple, as is the processing of the results to yield the calibration parameters.

10.4 Summary of conclusions

The problem of capturing stereoscopic images suitable for human viewing has been partially addressed by software calibrations, yet such calibrations have many drawbacks. They may be unstable and require frequent recalibration, the calibration may only be valid at certain distances, the determination of camera parameters may rely on a complex and unreliable optimisation algorithm, and the rectification necessary causes loss of image size and resolution. A hardware-calibrated camera avoids such drawbacks, but is difficult to create, largely because of the difficulty in decoupling the effects of rotation and translation, which cause a camera to only perform at its best at one scene Z-distance.

The concept of separating the lens and the sensor in initial alignment (discussed in section 3.2) forms the basis of a solution to the problem of creating a hardware-calibrated camera. The

calculation of errors due to rotation, translation, principal point and principal distance variation (in section 4) quantifies the relationship between human requirements and camera attributes. The procedures devised and shown (in sections 5 to 7) for aligning, mounting and calibrating the camera system provide a blueprint for making such a camera, which is capable of producing images suitable for human viewing.

Software calibrations form the bulk of those commonly used in modern stereoscopy, and at the very least the cameras they use should take into account the mounting and instability issues raised in this thesis for maximum accuracy. The study of image translation with lens focusing shows how without accounting for these translations, conventional calibrations restrict their usefulness to the focal positions at which the camera pair was calibrated. For applications where preserving as much of the image resolution as possible is important, the laser alignment procedure and focal scale calibration shows how rectification which necessarily blurs images through interpolation can be avoided.

A hardware-based calibration avoids many of the problems associated with most stereoscopic cameras, and has been shown to produce stereoscopic images ready for display with a minimum of processing. A camera calibrated in hardware produces viewable images that retain as much image information as possible, information that is degraded by the rectification process associated with software calibrations. The information lost due to poor camera alignment in a stereo image pair can never be fully recovered in software, and for this reason a hardware calibration should be considered wherever maximum useable image size and resolution are important.

10.5 Future extension of the work in this thesis

10.5.1 Lens system stability

If lenses can be made which do not suffer optical centre shifts under forces applied to their mountings, a lens calibration as described in section 6 would suffer smaller errors, and would be more strongly valid over long timescales. Investigations of lens systems which achieve this would benefit the stability of stereoscopic camera calibrations.

Instead of using the lenses' built in focusing mechanisms, simple translation of the lens relative to the camera may provide a more stable method of focusing the images. This presents difficulties in terms of keeping dust and other contaminants from the interior of the camera, and also from any translation apparatus used, but these difficulties might be productively solved.

Finally, the use of kinematic mountings may allow for the highest level of lens position repeatability. This would require significant mechanical changes to conventional lens-camera systems, but if extended to kinematically mount all lens elements relative to the camera body, would completely solve the problem of lens system instability.

10.5.2 Lens distortion

Reduction of the effects of distortion in hardware must be achieved if stereo cameras are to achieve subpixel false vertical disparities for a wide range of scenes. Two schemes are suggested for proceeding in this direction, described below.

With a standard distortion correction as described in section 9.1, the total vertical disparity would be the combination of the errors from the alignment procedure, the lens calibration and a lens distortion calibration. The error from the latter should be significantly smaller than 0.75 pixels (which is the upper limit of the total errors for all camera parameters in software calibrations[74,110]), for a total vertical disparity of within ±0.90 pixels. This would meet the requirements set by Panum's area under the conditions set out in section 1.6.6, and would exceed the accuracy of a calibration which does not account for lens distortion. This treatment could be applied to the camera described in this thesis without any alteration to the hardware, but requires an interpolative correction to the images.

With the yaw correction for simple radial distortion, the total vertical disparity would be the combination of the errors from the alignment procedure, the lens calibration and the yaw correction. The error from the latter is estimated to be approximately ±1.86 pixels from section 9.3 for a scene range from 1m to infinity, for a total vertical disparity of within ±2.35 pixels. This would meet the requirements set by Panum's area under the conditions set out in section 1.6.6, and would exceed the accuracy of the calibration which does not account for lens distortion. This correction would not require any interpolative correction, but would require the cameras to be aligned with a divergent yaw.

10.6 Future applications of the work in this thesis

10.6.1 Use of digital zoom

A digital zoom function is sometimes used to assist or replace the optical zoom function of a camera. Instead of adjusting the focal length of the lens system to achieve a smaller field of view, the field of view is adjusted through selecting part of a digital image and enlarging it to the full frame size.

Digital zoom is generally not as useful as optical zoom, because the resolution of a digitally zoomed image is reduced in proportion to the zoom factor, while optically zoomed images retain full sensor resolution. However, digital zooms can be useful in certain circumstances, for example where moving parts in the camera are not permitted, or where part of an imaged scene must be zoomed in on after the image has been taken.

Because the calibration scheme implemented in this thesis is non-interpolative, it maintains full sensor resolution in the stereoscopic image pair. This resolution difference is relatively small when viewing images containing several million pixels, but becomes more significant as the image is digitally zoomed. Stereoscopic images which have been scaled in this way will have a higher scene resolution when a hardware calibration (as in this thesis) has been used, as opposed to a software calibration.

While highly digitally zoomed stereoscopic images are unlikely to find popular uses in everyday photography due to the low display resolution, the technique may find important applications in areas such as defence or security. Stereoscopic camera systems in these areas may be used to capture digital images for later stereoscopic inspection, and digital zoom may play a part in inspection of details in small areas of the image.

10.6.2 Mass production of calibrated cameras

Stereoscopic cameras have a whole range of possible applications, from professional applications where hardware may be built to order or manufactured in low volume, to consumer applications where costs are low and volumes are high. In order to manufacture high volumes of stereoscopic cameras, it is likely that fast, reliable, automated camera calibration would be required.

The laser alignment method used in this thesis is very well suited to fast automation. Because the only measurements required are two two-dimensional spot positions, detectors such as position sensitive detectors can provide real time feedback to actuators controlling the sensor alignment. This raises the possibility of calibrating a large number of cameras in a given time period using a small number of calibration rigs, and this may in future help facilitate the manufacture of stereoscopic cameras for the mass market.

10.6.3 Stereoscopic photography

Because the focus of this thesis is on creating a camera to capture images for human stereoscopic viewing rather than for three-dimensional scene analysis, the images produced are well-suited to being stereoscopically displayed. Techniques such as the offsetting of radial lens distortion through camera yaw are specifically designed to minimise vertical rather than horizontal disparity, and so are better suited to comfortable human viewing than to scene depth measurements.

Human stereoscopic image viewing is currently much less common than human monoscopic image viewing. While monoscopic viewing can be a flexible and powerful technique for communicating visual information, stereoscopic viewing is superior when presenting images of complex three-dimensional scenes, especially when foreground and background are hard to distinguish monoscopically. However, stereoscopic images may be uncomfortable to view if vertical disparity is not minimised, and so the techniques described in this thesis should be considered when designing stereoscopic imaging systems for this purpose.

10.6.4 Telepresence

Telepresence is an effect where a person experiences the world from a location which does not coincide with their physical location, and allows a person to participate in activities from a distance. Achieving telepresence depends on capturing, transmitting and recreating data for all of the human senses, and is generally a matter of degree, as no system can currently achieve this perfectly for all senses.

Replicating a person's visual system requires the use of stereoscopy. Using a hardware-calibrated stereoscopic camera to capture the visual information allows the display of that information with a minimum of image processing, reducing the time difference between the virtual and physical viewpoints. As such, use of the techniques developed in this thesis for developing telepresence systems would be of benefit.

10.6.5 Medical imaging

Medical imaging is often at the forefront of imaging technology. It is often valuable to obtain information from a variety of viewpoints, at a variety of wavelengths, and present this information to medical personnel in order to aid diagnosis or treatment. Such images are frequently complex, and may lack clear distinctions between image foreground and background, and these distinctions may be made more apparent through the use of stereoscopy.

Stereoscopic video may also be used to enhance remote surgery, where a surgeon performs an operation at a large distance from the patient, driving a robot which performs the actual surgery. In order for such procedures to be carried out well, the surgeon must be able to view the subject in enough detail to successfully direct the apparatus. Stereoscopic imaging may assist in enhancing the perception of depth in the surgeon's images, but the images must be properly presented for the surgeon's comfortable viewing. This relies on camera calibration, and the calibration described in this thesis is highly suitable for this purpose.

Because the laser alignment procedure described in chapter 5 is not based upon forming an image, the method does not depend on the wavelength at which a camera operates. The calibration technique is the same for any sensor, and the accuracy of the rotational calibration is not limited by the sensor resolution in the same way as, for example, conventional calibrations of low-resolution infrared image sensors.

10.6.6 Remote operation of vehicles

It is sometimes desirable to operate vehicles remotely. Usually this stems from the environment in which the vehicle operates, which may be too inhospitable, dangerous, small or remote to make use of an in situ human operator. Examples of areas where ROVs (Remotely Operated Vehicles) may be used include nuclear reactors, bomb disposal, excavations and planetary exploration. Stereoscopic imaging (compared to monoscopic imaging) allows the remote operator to more accurately assess the environment through which the vehicle is travelling, and this allows the operator to control the vehicle more appropriately.

The small amount of post-processing required with a hardware calibrated stereoscopic camera means that images can be displayed more quickly than with software calibrations, reducing the time delay that the operator must cope with.

10.6.7 Stereoscopic image matching and scene analysis

In stereoscopically analysing a scene, corresponding image features are normally required to be extracted. To aid this process, stereoscopic images are often rectified, warping the images so that corresponding points have no vertical disparity. This narrows the search space for matching points from the whole image to a single line, and allows matching algorithms to operate more quickly and with fewer errors.

Because the calibration in this thesis is designed to minimise vertical disparity, it may be used in place of rectification algorithms to optimise stereoscopic images for scene analysis. The error analysis in this thesis defines the search space over which any matching algorithm must operate, and the stable mounting means that a single calibration can determine this search space for all future images.

This would make three-dimensional scene analysis from stereoscopic images more simple and robust.

11 Appendices

The following appendices give summaries of the laser alignment procedure, lens calibration procedure and point extraction algorithm. The summaries are not a substitute for sections 5 and 6, and are intended as a practical guide for people who have read the relevant chapters and wish to use the methods therein to align and calibrate a stereoscopic camera pair.

11.1 Laser alignment procedure

1	Assemble the equipment: optical bench, raised bench, camera rail, 90° L-piece, laser rail, laser, attenuating filter, laser rail slider with pinhole screen, camera mountings, 6-axis mount (example in Figure 50), optical post with small height-adjustable screen, screen.
2	Set the equipment up as shown in Figure 51
3	Warm the laser up for a minimum of 30 minutes
4	Align the laser to the laser rail, using the laser rail slider with screen, as in Figure 48
5	Align the camera rail at 90° to the laser rail using the 90° L-piece, as shown in Figure 49
6	Attach the left camera to the 6-axis mount
7	Move the left camera to align the attenuated laser spot with the CCD centre
8	Align the left camera approximately in roll, pitch and yaw using laser rail slider with pinhole screen
9	Align the left camera in roll and pitch by sliding the post with screen around the bench, adjusting roll and pitch to give horizontally-diffracted beams in a plane parallel to the bench surface
10	Fix crosshairs on the screen, aligned with a pair of horizontally widely diffracted points
11	Fix the left camera in camera mounting
12	Remove the 6-axis mount from the left camera, translate the camera rail to the left, slide the 6-axis mount to the right camera position
13	Attach the right camera to the 6-axis mount
14	Move the right camera to align the attenuated laser spot with the CCD centre
15	Align the right camera in roll, pitch, yaw and Z by aligning the horizontally diffracted points with the crosshairs on the screen
16	Align the right camera in Y using callipers, with an equal camera base to camera rail spacing to that of the left camera

17	Fix the right camera in the camera mounting
18	Remove the 6-axis mount

11.2 Lens calibration procedure

1	Assemble the equipment: laser-aligned camera pair, connection to PC, pair of lenses, backlit calibration target with Gaussian blurred spots (to fill most of the overlapping fields of view of the cameras), apparatus to read lens focus barrel rotations.
2	Fit the lenses to the cameras and fix the camera rail in position
3	Position the target parallel to the camera rail, with the axis approximately aligned with an axis midway between the cameras' principal axes
4	Take one image with each camera at sharpest focus (with maximum aperture)
5	Process the images and extract the point sets
6	Calculate the image scale of each point set
7	Check to make sure the scales are within 0.24% of each other, if they are not, choose more closely matched lenses and start again from step 1
8	Set minimum aperture on the lenses, high exposure time for the cameras, and adjust the camera gains and offsets to give good contrast between calibration points and background if necessary
9	Take images with the left camera across the focal range to be calibrated, at regular angular intervals, repeating for n data sets (higher n demonstrates higher repeatability)
10	Take images with the right camera across the focal range to be calibrated, at the same regular angular intervals as for the left camera, repeating for n data sets
11	Process the images and extract the point sets
12	Calculate the centre of mass, and the mean point-to-point distance (a measure of the image scale) for every point set
13	Calculate the mean centre of mass for each camera's point set at each focal position, averaging over the n sets
14	Fit a linear function to the total scale data across all results for each camera, of the form scale = a + b * focus position
15	Use Equation 47 to calculate the right camera focus position as a function of left camera focus position
16	For the range of measured left camera focus positions, calculate the right camera centre of mass values by linear interpolation
17	Calculate the predicted image translations as the difference between the left camera centre of mass and the interpolated right camera centre of mass, for left camera focus positions and corresponding calculated (for equal scale) right camera focus positions.

	Apply a correction for the effects of lens distortion, from modelling as described in section 6.1.10

11.3 Point extraction procedure

1	Colour interpolate each image if necessary
2	Set the initial threshold estimate, calibration point minimum separation value (will depend on size of calibration grid as seen by the camera), number of calibration points, fraction of threshold to centroid at (will depend on target contrast, lighting, camera response), centroid window size (will depend on the size of the points as imaged), and maximum point movement between images (will depend on target, lens stability and focus step size)
3	Threshold the first image, rejecting any point above the threshold closer than the minimum separation value to a previously detected point, unless the pixel value is higher than that of the previously detected point, in which case the old point is overwritten with the new one.
4	Count the thresholded points, if the number is higher than the number of calibration points then the threshold is increased and the process repeats from step 3, if the number is lower than the number of calibration points then the threshold is decreased and the process repeats from step 3, and if the number is equal to the number of calibration points then the process advances to step 5
5	Each thresholded point should be centroided with the defined window size, with all pixels having values above the defined centroiding fraction multiplied by the threshold contributing to the centroid sum
6	Steps 3 to 5 must be repeated for every image in the data set
7	Points must be validated across images in the data set, any point without a corresponding (within the maximum point movement value) detected point in every image in the set being discarded
8	The validated points should be recorded as the calibration point locations for each image, if there are a significant number of discarded points then results may be improved by manually adjusting initial parameter estimates to better suit the properties of the calibration images
9	Steps 3 to 8 are repeated for each data set, and for each camera

12 References

[1] "The Concise Oxford Dictionary", Fifth Edition, Oxford University Press, 1964

[2] "Stereo Photography – An Introduction To Stereo Photo Technology And Practical Suggestions For Stereo Photography", F. G. Waack, German Stereoscopic Society, 1985

[3] "The Photographer's Handbook 3rd Edition", J. Hedgecoe, Knopf, 1992

[4] "Constructing the Cyclopean View", R.D. Henkel, Proceedings of the International Conference on Artifical Neural Networks, pp907-912, 1997

[5] "The Hutchinson Encyclopaedia", Helicon Publishing, 2001

[6] "3-D or not 3-D?", C. Smith, New Scientist, 26 April 1984

[7] "Time, space and interaction", A. Dix, Proceedings of FADIVA 3, University of Rome, pp99-103, 1996.

[8] "Stereo-Realist Manual", Morgan & Morgan Inc., 1954

[9] "Design of a single-lens stereo camera system', A. Goshtasby and W. A. Gruver, Pattern Recognition, 26(6), pp923-937, 1993

[10] "Optical Holography", 2nd edition, P Hariharan, Cambridge University Press, 1996

[11] "Solid state volumetric display projecting 3D objects in space", E. van Nuland, Stereoscopic displays and applications XII, SPIE vol. 4297, 2001
pp236-250

[12] "Magic Eye: A New Way Of Looking At The World", N. E. Thing Enterprises, Magic Eye Inc., Andrews McNeel, 1993

[13] "New Autostereoscopic Display System", D. Ezra, G. J. Woodgate, B. A. Omar, N. S. Holliman, J. Harrold and L. S. Shapiro, SPIE vol. 2409, pp31-40, 1995.

[14] 3D imaging department, Sharp Laboratories of Europe

[15] "Performance of a Convertible 2D and 3D Parallax Barrier Autostereoscopic Display", J. Harrold, A. Jacobs, G. Woodgate and D. Ezra, Proc. SID, 20th International Display Research Conference, 2000.

[16] "Messiah", Shiny Entertainment Inc., 2000

[17] "Improving the Operability of Remotely Operated Vehicles", A. J. Woods, J. D. Penrose, A. J. Duncan, R. Koch and D. Clark, APPEA 98 conference, Canberra, March 1998.

[18] "The analysis of stereopsis", G. F. Poggio and T. Poggio, Annual Review of Neuroscience, vol. 7, pp379-412, 1984

[19] "Neural mechanisms underlying stereoscopic vision", F. Gonzalez and R. Perez, Progress in Neurobiology, 55 (3) pp191-224, June 1998

[20] "Extension of Panum's Fusional Area in Binocularly Stabilised Vision", D. Fender and B. Julesz, Journal of the Optical Society of America, vol. 57 no. 6, pp819-830, June 1967

[21] "The Human Nervous System, Structure And Function", 5th Edition, C. R. Noback, N. L. Strominger and R. J. Demarest, Williams & Wilkins, 1996

[22] "Enhancement of viewer comfort in stereoscopic viewing: parallax adjustment", J. Konrad, SPIE 3639 pp179-190, January 1999

[23] "The world of 3-D, A practical guide to stereo photography", J. G. Ferwerda, Netherlands Society for Stereo Photography, 1982

[24] "Human Engineering in Stereoscopic Viewing Devices", D. B. Diner and D. H. Fender, Plenum Press, 1993.

[25] "An Analysis of Different Methods of Acquiring Real World Stereoscopic Images: A Summary of Stereoscopic Camera Designs", D.J. Montgomery, C.K. Jones, J. N. Stewart and A. Smith, Stereoscopic Displays and Applications XIII, SPIE vol. 4660, 2002

[26] "Digital camera self-calibration", C. S. Fraser, ISPRS Journal of Photogrammetry and Remote Sensing, 52, pp149-159, 1997

[27] "Digital Stereoscopic Imaging", A. R. Rao and A. Jaimes, Stereoscopic Displays and Virtual Reality Systems VI, SPIE vol. 3639, pp144-154, 1999

[28] "Basler A101 User's Manual", Basler Vision Technologies, DA 037202, 10 July 2001

[29] "Kodak CCD Primer #KCP-001", Eastman Kodak Company Microelectronics Technology Division, DS00-003, rev. 0, August 1999

[30] "CNG 1.4/8 data sheet", Schneider Kreuznach, 0302

[31] "Hyperterminal", Hilgraeve Inc.

[32] "Common Vision Blox Image Manager", Stemmer Imaging GMBH

[33] "John Henshall's Chip Shop", J. Henshall, The Photographer, January 1993

[34] "Astrophysical Techniques", 2nd Edition, C. R. Kitchin, IOP Publishing Ltd., 1991

[35] "Solid State Image Sensors: Terminology", Kodak Application Note, DS 00-001

[36] "CMOS sensor systems", B. Hosticka, Sensors and Actuators A, 66, pp335-341, 1998

[37] "Electronic Imaging For Photographers, 2nd Edition", A. Davies and P. Fennessy, Focal Press, 1998

[38] "An introduction to CCD operation", CCD group, Mullard Space Science Laboratory, University College London

[39] "The advantages and disadvantages of small pixels", R. Palum, IS&T Proc. PICS 2000, March 2000

[40] "ICX085AK datasheet", Sony Electronics

[41] "Color Imaging Array", B. E. Bayer, US Patent 3971065

[42] "True Two-Phase CCD Image Sensors Employing a Transparent Gate", Kodak Application Note, DS 02-006

[43] "A practical guide to CCD astronomy", P. Martinez and A Klotz, Cambridge University Press, 1998

[44] "Handbook of CCD Astronomy", S. B. Howell, Cambridge University Press, 2000

[45] "Interline Image Sensor: KAI-1003M", Kodak Application Note, DS 02-027

[46] "The development of camera calibration methods and models", T. A. Clarke and J. G. Fryer, Photogrammetric Record, 16(91), pp51-66, 1998.

[47] "The principal point and CCD cameras", T. A. Clarke, J. G. Fryer and X. Wang, Photogrammetric Record, 16(92), pp293-312, 1998

[48] "When should we consider lens distortion in camera calibration", S. Shih, Y. Hung and W. Lin, Pattern Recognition, 28(3), pp447-461, 1995

[49] "Correction of image deformation from lens distortion using Bezier patches", A. Goshtasby, Computer Vision, Graphics and Image Processing, 47, pp385-394, 1989

[50] "Calibration Procedure for Short Focal Length Off-the-Shelf CCD Cameras", J. Heikkilä and Olli Silvén, IEEE Proceedings of the International Conference on Pattern Recognition, pp166-170, 1996.

[51] "A camera calibration technique using three sets of parallel lines", T. Echigo, Machine Vision and Applications, 3, pp 159-167, 1990

[52] "Camera self-calibration: theory and experiments", O. D. Faugeras, Q. Luong and S. J. Maybank, European Conference on Computer Vision, 1992

[53] "Digital camera self-calibration", C. S. Fraser, ISPRS Journal of Photogrammetry and Remote Sensing, 52, ppl49-159, 1997

[54] "CCD camera calibration based on natural landmarks", G. Zhou, U. Uzi, W. Feng and B. Yuan, Pattern Recognition, 31 (11), pp 1715-1724, 1998

[55] "Autocalibration from planar scenes", B. Triggs, European Conference on Computer Vision, 1998

[56] "Self-calibration from image triplets", M. Armstrong, A. Zisserman and R. Hartley, European Conference on Computer Vision, 1996

[57] "Do we really need an accurate calibration pattern to achieve a reliable camera calibration?", J. Lavest, M. Viala and M. Dhome, European Conference on Computer Vision, 1998

[58] "Camera calibration problems: some new results", R. J. Holt and A. N. Netravali, Computer Vision, Graphics and Image Processing: Image Understanding, 54(3), pp368-383, 1991

[59] "Camera Calibration with Distortion Models and Accuracy Evaluation", J. Weng, P. Cohen and M. Herniou, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 10, pp965-980, 1992.

[60] "A theory of self-calibration of a moving camera", S. J. Maybank and O. Faugeras, International Journal of Computer Vision, vol. 8, no. 2, pp123-151, 1992.

[61] "Self-calibration from multiple views with a rotating camera", R. Hartley, Eurpoean Conference on Computer Vision, pp471-478, 1994.

[62] "Decomposition of transformation matrices for robot vision", S. Ganapathy, Proceedings of the IEEE Conference on Robotics and Automation, pp130-139, 1984

[63] "Metric rectification for perspective images of planes", D. Lieberwitz and A. Zisserman, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp482-488, 1998.

[64] "Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology", R. K. Lenz and R. Y. Tsai, Proceedings of the IEEE Conference on Robotics and Automation, pp68-75, 1987.

[65] "Stereoscopic imaging via rotation and translation", M. A. Weissman, Stereoscopic Displays and Virtual Reality Systems II, SPIE vol. 2409, pp62-66, 1995.

[66] "Image Pickup Device, Stereoscopic Broadcast System and Stereoscopic Video System", A. Mochizuki et al, Japanese patent JP6197379,1994.

[67] "Imager for Mars Pathfinder experiment (IMP): a multispectral stereo imaging system", P. H. Smith, Stereoscopic Displays and Virtual Reality Systems V, SPIE vol. 3295, pp4-11, 1998.

[68] "Accurate and simple geometric calibration of multi-camera systems", F. Pedersini, A. Sarti and S. Tubaro, Signal Processing, vol. 77, pp309-334, 1999.

[69] "Image distortions in stereoscopic video displays", A. Woods, T. Docherty and R. Koch, Stereoscopic Displays and Applications, SPIE vol. 1915, 1993

[70] "Controlling perceived depth in stereoscopic images", G. R. Jones, D. Lee, N. S. Holliman, D. Ezra, Stereoscopic Displays and Virtual Reality Systems VIII, SPIE vol. 4297A, 2001

[71] "Principles of Optics : Electromagnetic Theory of Propagation, Interference and Diffraction of Light, 7th Edition", M. Born and E. Wolf, Cambridge University Press, 1999

[72] "Optics Source Book", S. P. Parker, McGraw-Hill, 1988

[73] "Computing Rectifying Homographies for Stereo Vision", C. Loop and Z. Zhang, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp125-131, 1999.

[74] "An Effective Technique for Calibrating a Binocular Stereo Through Projective Reconstruction Using Both a Calibration Object and the Environment", Z. Zhang, O. Faugeras and R. Deriche, Videre, vol. 1 (1), 1997.

[75] "Computing projective distortion matrix (collineation)", Z. Zhang, Technical report, INRIA Robotvis, 1993.

[76] "The calibration problem for stereo", O. Faugeras and G. Toscani, Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp15-20, 1986.

[77] "What can be seen in three dimensions with an uncalibrated stereo rig", O. Faugeras, Proceedings of the 2nd European Conference on Computer Vision, 1992.

[78] "Stereo from uncalibrated cameras", R. Hartley, R. Gupta and T. Chang, Proceddings of the IEEE Conference on Computer Vision and Pattern Recognition, pp761-764, 1992.

[79] "A robust technique for matching two uncalibrated images through the recovery of unknown epipolar geometry", Z. Zhang, R. Deriche, O. Faugeras and Q. T. Luong, Artificial Intelligence Journal, vol. 78, pp87-119, 1995.

[80] "Closed-Form Solutions for the Euclidean Calibration of a Stereo Rig", G. Csurka, D. Demirdijan, A. Ruf and R. Horaud, European Conference on Computer Vision, pp426-442, 1998.

[81] "Rectification with unconstrained stereo geometry", A. Fusiello, E. Trucco and A. Verri, BMVC97, pp400-409, 1997.

[82] "Mounting Lenses In Optical Instruments", P. R. Yoder Jr., Tutorial Texts in Optical Engineering, vol. TT21, SPIE Optical Engineering Press, 1995.

[83] "Effects of unequal focal lengths in stereo imaging", S. Sengupta, Pattern Recognition Letters, 18, pp395-400, 1997

[84] "Effects of camera alignment errors on stereoscopic depth estimates", W. Zhao and N. Nandhakumar, Pattern Recognition, 29 (12), pp2115-2126, 1996

[85] "A Flexible New Technique for Camera Calibration", Z. Zhang, Microsoft Technical Report MSR-TR-98-71, March 1999

[86] "Introductory Techniques for 3-D Computer Vision", E. Trucco and A. Verri, Prentice Hall, pp332-333, 1998

[87] "The Duality and Critical Condition in the Formulation and Decomposition of a Rotation Matrix", T. Y. Shih, Photogrammetric Engineering and Remote Sensing, 56(8), pp1173-1179

[88] "Imatrig user manual", Imasys

[89] "University Physics", H. Benson, John Wiley & Sons, 1991

[90] "PRL series documentation", Newport Corporation

[91] "Optical Rails and Carriers", Melles Griot

[92] "PRC-3 specifications", Newport Corporation

[93] "ICX085AL datasheet", Sony Electronics

[94] "B+W 486 filter specifications", Schneider Optics

[95] "423/433/443 series documentation", Newport Corporation

[96] "A Progressive Scheme for Stereo Matching", Z. Zhang and Y. Shan, 3D Structure from Images - SMILE 2000, LNCS 2018, pp68-85, 2001.

[97] "Machine Vision: Theory, Algorithms, Practicalities, 2nd Edition", E. R. Davies, Academic Press, 1997.

[98] "A comparison of digital centering algorithms", R. C. Stone, Astronomical Journal, vol. 97, pp1227-1237, 1989.

[99] "Precision CCD Astronomy", B. Mendez and P. O. Seitzer, Bulletin of the American Astronomical Society, vol. 28, 1996.

[100] "Determination of Optical Positions for Extragalactic Radio Sources under the Collaboration between SHAO and NAO", Z. Tang, W. Jin, S. Wang, G. Pinigin, A. Shulga, N. Magurova and Y. Protsyuk, Proceeding of the IEEE Colloquium 180, 2000.

[101] "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision", R. Y. Tsai, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp364-374, 1986.

[102] "Paint Shop Pro 6.0", Jasc Software

[103] "Inside Windows File Formats", T. Swan, Sams Publishing, 1993

[104] PBMPlus toolkit, Jeff Poskanzer, ACME Labs Software

[105] "Microsoft Excel 97", Microsoft Corporation

[106] "Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity and Panmixia", K. Pearson, Philosophical Transaction of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character, vol. 187, pp253-318, 1896.

[107] "Scotchweld 1838 instructions", 3M

[108] A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", D. Scharstein and R. Szeliski, International Journal of Computer Vision, 47 (1), pp7-42, 2002.

[109] "A Camera Calibration Method Using Parallelogrammetric Grid Points", A. Takahashi, I.Ishii, H. Makino and M. Nakashizuka, IEICE Transactions on Information and Systems, vol. E79-D (11), 1996.

[110] "Camera calibration from multiple views of a 2D object, using a global non linear minimization method", M. Devy, V. Garric and J. J. Orteu, Proc. IROS 97, 1997.

[111] "Handbook of Mathematical Functions, With Formulas, Graphs and Mathematical Tables", M. Abramowitz and I. A. Stegun, Dover Publications, 1974.